

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
18 August 2005 (18.08.2005)

PCT

(10) International Publication Number
WO 2005/076010 A2

(51) International Patent Classification⁷: **G01N 33/68**

(21) International Application Number:
PCT/IN2005/000037

(22) International Filing Date: 7 February 2005 (07.02.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
173/DEL/2004 6 February 2004 (06.02.2004) IN
60/589,227 20 July 2004 (20.07.2004) US

(71) Applicant (for all designated States except US): **COUNCIL OF SCIENTIFIC AND INDUSTRIAL RESEARCH** [IN/IN]; Rafi Marg, 110 001 New Delhi, Maharashtra (IN).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **SACHDEVA, Gaurav** [IN/IN]; Institute of Genomics and Integrative Biology, Mall Road, Dehli 110 007 (IN). **KUMAR, Kaushal** [IN/IN]; Institute of Genomics and Integrative Biology, Mall Road, Dehli 110 007 (IN). **JAIN, Preti** [IN/IN]; Institute of Genomics and Integrative Biology, Mall Road, Dehli 110 007 (IN). **BRAHMACHARI, Samir, Kumar** [IN/IN]; Institute of Genomics and Integrative Biology, Mall Road, New Delhi 110 007 (IN). **RAMACHANDRAN, Srinivasan** [IN/IN]; Institute of Genomics and Integrative Biology, Mall Road, Dehli 110 007 (IN).

(74) Agents: **BHOLA, Ravi** et al.; K & S Partners, 84-C, C6 Lane, Off Central Avenue, Sainik Farms, New Delhi 110 067 (IN).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for all designations
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for all designations

Published:

- without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE PROTEINS OF THERAPEUTIC POTENTIAL

(57) Abstract: A computational method for identifying adhesin and adhesin-like proteins, said method comprising steps of computing the sequence-based attributes of a neural network software wherein the attributes are (i) amino acid frequencies, (ii) multiplet frequency, (iii) dipeptide frequencies, (iv) charge composition, and (v) hydrophobic composition, training the artificial neural Network (ANN) for each of the computed five attributes, and identifying the adhesin and adhesin-like proteins having probability of being an adhesin (P_{ad}) as ≥ 0.51 ; a computer system for performing the method; and genes and proteins encoding adhesin and adhesin-like proteins.



WO 2005/076010 A2

COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE PROTEINS OF THERAPEUTIC POTENTIAL

Field of the present Invention

5 A computational method for identifying adhesin and adhesin-like proteins; computer system for performing the method; and genes and proteins encoding adhesin and adhesin-like proteins

Background and prior art of the present invention

10 The progress in genome sequencing projects has generated a large number of inferred protein sequences from different organisms. It is expected that the availability of the information on the complete set of proteins from infectious human pathogens will enable us to develop novel molecular approaches to combat them. A necessary step in the successful colonization and subsequent manifestation of disease by microbial pathogens is the ability to adhere to host cells.

15 Microbial pathogens encode several proteins known as adhesins that mediate their adherence to host cell surface receptors, membranes, or extracellular matrix for successful colonization. Investigations in this primary event of host-pathogen interaction over the past decades have revealed a wide array of adhesins in a variety of pathogenic microbes. Presently, substantial information on the biogenesis of adhesins and the regulation of adhesin factors is available. One of the best understood mechanisms of bacterial adherence is attachment mediated by pili or fimbriae. Several afimbrial adhesins also have been reported. In addition, limited knowledge on the target host receptors also has been gained (Finlay, B.B. and Falkow, S 1997).

25 New approaches to vaccine development focus on targeting adhesins to abrogate the colonization process (Wizemann, *et al* 1999). However, the specific role of particular adhesins has been difficult to elucidate. Thus, prediction of adhesins or adhesin-like proteins and their functional characterization is likely to aid not only in deciphering the molecular mechanisms of host pathogen interaction but also in developing new vaccine formulations, which can be tested in suitable experimental model systems.

30 One of the best understood mechanisms of bacterial adherence is attachment mediated by pili or fimbriae. For example, FimH and PapG adhesins of *Escherichia coli* (Maurer, L., Orndorff, P.(1987), Bock, K., *et al.*(1985). Other examples of pili group adhesins include type IV pili in *Pseudomonas aeruginosa*, *Neisseria* species, *Moraxella* species, Enteropathogenic *Escherichia coli* and *Vibrio cholerae* (Sperandio V *et al* (1996).

Several afimbrial adhesins are HMW proteins of *Haemophilus influenzae* (van Schilfgaarde 2000), the filamentous hemagglutinin, pertactin, of *Bordetella pertussis* (Bassinet *et al* 2000), the BabA of *H. pylori* (Yu J *et al* 2002) and the YadA adhesin of *Yersinia enterocolitica* (Neubauer *et al* 2000). The intimin receptor protein (Tir) of Enteropathogenic *E. coli* (EPEC) is another type of adhesin (Ide T *et al* 2003). Other class of adhesins includes MrkD protein of *Klebsella pneumoniae*, Hia of *H. influenzae* (St Geme *et al* 2000), Ag I/II of *Streptococcus mutans* and SspA, SspB of *Streptococcus gordonii* (Egland *et al* 2001), FnbA, FnbB of *Staphylococcus aureus* and SfbI, protein F of *Streptococcus pyogenes*, the PsaA of *Streptococcus pneumoniae* (De
5 10 *et al* 2003).

A known example of adhesins approved as vaccine is the acellular pertussis vaccine containing FHA and pertactin against *B. pertussis* the causative agent of whooping cough (Halperin, S *et al* 2003). Immunization with FimH is being evaluated for protective immunity against pathogenic *E. coli* (Langermann S *et al* 2000), in
15 *Streptococcus pneumoniae*, PsaA is being investigated as a potential vaccine candidate against pneumococcal disease (Rapola, S *et al* 2003). Immunization results with BabA adhesin showed promise for developing a vaccine against *H. pylori* (Prinz, C *et al* 2003). A synthetic peptide sequence anti-adhesin vaccine is being evaluated for protection against *Pseudomonas aeruginosa* infections.

20 Screening for adhesin and adhesin like proteins by conventional experimental method is laborious, time consuming and expensive. As an alternative, homology search is used to facilitate the identification of adhesins. Although, this procedure is useful in the analysis of genome organization (Wolf *et al* 2001) and of metabolic pathways (Peregrin-Alvarez *et al* 2003, Rison *et al* 2002), it is somewhat limited in allowing
25 functional predictions when the homologues are not functionally characterized or the sequence divergence is high. Assignment of functional roles to proteins based on this technique has been possible for only about 60% of the predicted protein sequences (Fraser *et al* 2000). Thus, we explored the possibility of developing a non-homology method based on sequence composition properties combined with the power of the
30 Artificial Neural Networks to identify adhesins and adhesin-like proteins in species belonging to wide phylogenetic spectrum.

Twenty years ago, Nishikawa *et al* carried out some of the early attempts to classify proteins into different groups based on compositional analysis (Nishikawa *et al* 1983).

More recently, the software PropSearch was developed for analyzing protein sequences where conventional alignment tools fail to identify significantly similar sequences (Hobohm, U. and Sander, C 1995). PropSearch uses 144 compositional properties of protein sequences to detect possible structural or functional relationships between a new sequence and sequences in the database. Recently the compositional attributes of proteins have been used to develop softwares for predicting secretory proteins in bacteria and apicoplast targeted proteins in *Plasmodium falciparum* by training Artificial Neural Networks (Zuegge *et al* 2001).

Zuegge *et al* have used the 20 amino acid compositional properties. Their objective was to extract features of apicoplast targeted proteins in *Plasmodium falciparum*. This is distinct from our software SPAAN that focuses on adhesins and adhesin-like proteins involved in host-pathogen interaction.

Hobohm and Sander have used 144 compositional properties including isoelectric point and amino acid and dipeptide composition to generate hypotheses on putative functional role of proteins that are refractory to analysis using other sequence alignment based approaches like BLAST and FASTA. Hobohm and Sander do not specifically address the issue of adhesins and adhesin-like proteins, which is the focus of SPAAN

Nishikawa *et al* had originally attempted to classify proteins into various functional groups. This was a curiosity driven exercise but eventually lead to the development of a software to discriminate extra-cellular proteins from intracellular proteins. This work did not address the issue of adhesins and adhesin-like proteins, which is the focus of SPAAN.

Thus, none of the aforementioned research groups have been able to envisage the methodology of the instant application. The inventive method of this application provides novel proteins and corresponding gene sequences.

Adhesins and adhesin-like proteins mediate host-pathogen interactions. This is the first step in colonization of a host by microbial pathogens. Attempts Worldwide are focused on designing vaccine formulations comprising adhesin proteins derived from pathogens. When immunized, host will have its immune system primed against adhesins for that pathogen. When a pathogen is actually encountered, the surveillance mechanism will recognize these adhesins, bind them through antigen-antibody interactions and neutralize the pathogen through complement mediate cascade and other related clearance mechanisms. This strategy has been successfully employed in

the case of Whooping cough and is being actively pursued in the case of Pneumonia, Gastric Ulcer and Urinary tract infections.

Objects of the present Invention

The main object of the present invention is to provide a computational method for identifying adhesin and adhesin-like proteins of therapeutic potential.

Another object of invention is to provide a method for screening the proteins with unique compositional characteristics as putative adhesins in different pathogens.

Yet, another object of the invention is providing the use of gene sequences encoding the putative adhesin proteins useful as preventive therapeutics.

Summary of the present Invention

A computational method for identifying adhesin and adhesin-like proteins, said method comprising steps of computing the sequence-based attributes of protein sequences using five attribute modules of software SPAAN, (i) amino acid frequencies, (ii) multiplet frequency, (iii) dipeptide frequencies, (iv) charge composition, and (v) hydrophobic composition, training the artificial neural Network (ANN) for each of the computed five attributes, and identifying the adhesin and adhesin-like proteins having probability of being an adhesin (P_{ad}) as ≥ 0.51 ; a computer system for performing the method; and genes and proteins encoding adhesin and adhesin-like proteins

Detailed description of the present Invention

Accordingly, the present invention relates to a computational method for identifying adhesin and adhesin-like proteins, said method comprising steps of computing the sequence-based attributes of protein sequences using five attribute modules of software SPAAN, (i) amino acid frequencies, (ii) multiplet frequency, (iii) dipeptide frequencies, (iv) charge composition, and (v) hydrophobic composition, training the artificial neural Network (ANN) for each of the computed five attributes, and identifying the adhesin and adhesin-like proteins having probability of being an adhesin (P_{ad}) as ≥ 0.51 ; a computer system for performing the method; and genes and proteins encoding adhesin and adhesin-like proteins

In an embodiment of the present invention, wherein the invention relates to a computational method for identifying adhesin and adhesin-like proteins, said method comprising steps of:

- a. computing the sequence-based attributes of protein sequences using five attribute modules of a neural network software, wherein the attributes

are software, (i) amino acid frequencies, (ii) multiplet frequency, (iii) dipeptide frequencies, (iv) charge composition, and (v) hydrophobic composition,

- b. training the artificial neural Network (ANN) for each of the computed five attributes, and
- c. identifying the adhesin and adhesin-like proteins having probability of being an adhesin (P_{ad}) as ≥ 0.51 .

In another embodiment of the present invention, wherein the invention relates to a method wherein the protein sequences is obtained from pathogens, eukaryotes, and multicellular organisms.

In an embodiment of the present invention, wherein the invention relates to a method, wherein the protein sequences are obtained from the pathogens selected from a group of organisms comprising *Escherichia coli*, *Haemophilus influenzae*, *Helicobacter pylori*, *Mycoplasma pneumoniae*, *Mycobacterium tuberculosis*, *Rickettsiae prowazekii*, *Porphyromonas gingivalis*, *Shigella flexneri*, *Streptococcus mutans*, *Streptococcus pneumoniae*, *Neisseria meningitides*, *Streptococcus pyogenes*, *Treponema pallidum* and Severe Acute Respiratory Syndrome associated human coronavirus (SARS).

In yet another embodiment of the present invention, wherein the method of the invention is a non-homology method.

In still another embodiment of the present invention, wherein the invention relates to the method using 105 compositional properties of the sequences.

In still another embodiment of the present invention, wherein the invention relates to a method showing sensitivity of at least 90%.

In still another embodiment of the present invention, wherein the invention relates to the method showing specificity of 100%.

In still another embodiment of the present invention, wherein the invention relates to a method identifying adhesins from distantly related organisms.

In still another embodiment of the present invention, wherein the invention relates to the neural network has multi-layer feed forward topology, consisting of an input layer, one hidden layer, and an output layer.

In still another embodiment of the present invention, wherein the invention relates to the number of neurons in the input layer are equal to the number of input data points for each attribute.

In still another embodiment of the present invention, wherein the invention relates to the "P_{ad}" is a weighted linear sum of the probabilities from five computed attributes.

In still another embodiment of the present invention, wherein the invention relates to each trained network assigns a probability value of being an adhesin for the protein sequence.

5 In still another embodiment of the present invention, wherein the invention relates to a computer system for performing the method of claim 1, said system comprising a central processing unit, executing SPAAN program, giving probabilities based on different attributes using Artificial Neural Network and in built other programs of assessing attributes, all stored in a memory device accessed by CPU, a display on
10 which the central processing unit displays the screens of the above mentioned programs in response to user inputs; and a user interface device.

In still another embodiment of the present invention, wherein the invention relates to a set of 274 annotated genes encoding adhesin and adhesin-like proteins, having SEQ ID
15 Nos. 385 to 658.

In still another embodiment of the present invention, wherein the invention relates to a set of 105 hypothetical genes encoding adhesin and adhesin-like proteins, having SEQ ID Nos. 659 to 763.

In still another embodiment of the present invention, wherein the invention relates to a set of 279 annotated adhesin and adhesin-like proteins of SEQ ID Nos. 1 to 279.
20

In still another embodiment of the present invention, wherein the invention relates to a set of 105 hypothetical adhesin and adhesin-like proteins of SEQ ID Nos. 280 to 384.

One more embodiment of the present invention, wherein the invention also relates to a fully connected multilayer feed forward Artificial Neural Network based on the computational method as claimed in claim 1, comprising of an input layer, a hidden
25 layer and an output layer which are connected in the said sequence, wherein each neuron is a binary digit number and is connected to each neuron of the subsequent layer for identifying adhesin or adhesin like proteins, wherein the program steps comprise:-
[a] feeding a protein sequence in FASTA format; [b] processing the sequence
30 obtained in step [a] through the 5 modules named A, C, D, H and M, wherein attribute A represents an amino acid composition, attribute C represents a charge composition, attribute D represents a dipeptide composition of the 20 dipeptides [NG, RE, TN, NT, GT, TT, DE, ER, RR, RK, RI, AT, TS, IV, SG, GS, TG, GN, VI and HR], attribute H

represents a hydrophobic composition and attribute M represents amino acid frequencies in multiplets to quantify 5 types of compositional attributes of the said protein sequence to obtain numerical input vectors respectively for each of the said attributes wherein the sum of numerical input vectors is 105; [c] processing of the numerical input vectors obtained in step [b] by the input neuron layer to obtain signals, wherein the number of neurons is equal to the number of numerical input vectors for each attribute; [d] processing of signals obtained from step [c] by the hidden layer to obtain synaptic weighted signals, wherein the optimal number of neurons in the hidden layer was determined through experimentation for minimizing the error at the best epoch for each network individually; [e] delivering synaptic weighted signals obtained from step [d] to the output layer for assigning of a probability value for each protein sequence fed in step [a] as being an adhesin by each network module; [f] using the individual probabilities obtained from step [e] for computing the final probability of a protein sequence being an adhesin denoted by the P_{ad} value, which is a weighted average of the individual probabilities obtained from step [e] and the associated fraction of correlation which is a measure of the strength of the prediction.

In still another embodiment of the present invention, wherein the input neuron layer consists of a total of 105 neurons corresponding to 105 compositional properties.

In still another embodiment of the present invention, wherein the hidden layer comprises of neurons represented as 30 for amino acid frequencies, 28 for multiplet frequencies, 28 for dipeptide frequencies, 30 for charge composition and 30 for hydrophobic composition.

In still another embodiment of the present invention, wherein the output layer comprises of neurons to deliver the output values as probability value for each protein sequence.

Identification of novel adhesins and their characterization are important for studying host-pathogen interactions and testing new vaccine formulations. We have employed Artificial Neural Networks to develop an algorithm SPAAN (Software for Prediction of Adhesin and Adhesin-like proteins using Neural Networks) that can identify adhesin proteins using 105 compositional properties of a protein sequence. SPAAN could correctly predict well characterized adhesins from several bacterial species and strains. SPAAN showed 89% sensitivity and 100% specificity in a test data set that did not

contain proteins in the training set. Putative adhesins identified by the software can serve as potential preventive therapeutics.

The present invention provides a novel computational method for identifying adhesin and adhesin-like proteins of therapeutic potential. More particularly, the present invention relates to candidate genes for these adhesins. The invention further provides new leads for development of candidate genes, and their encoded proteins in their functional relevance to preventive approaches. This computational method involves calculation of several sequence attributes and their subsequent analyses lead to the identification of adhesin proteins in different pathogens. Thus, the present invention is useful for identification of the adhesin proteins in pathogenic organisms. The adhesin proteins from different genomes constitute a set of candidates for functional characterization through targeted gene disruption, microarrays and proteomics. Further, these proteins constitute a set of candidates for further testing in development of preventive therapeutics. Also, are provided the genes encoding the candidate adhesin proteins.

The present method offers novelty in the principles used and the power of Neural Networks to identify new adhesins compared to laborious and time consuming conventional methods. The present method is based on compositional properties of proteins instead of sequence alignments. Therefore this method has the ability to identify adhesin and adhesin like proteins from bacteria belonging to a wide phylogenetic spectrum. The predictions made from this method are readily verifiable through independent analysis and experimentation. The invention has the potential to accelerate the development of new preventive therapeutics, which currently requires high investment in terms of requirement of skilled labor and valuable time.

The present invention relates to a computational method for the identification of candidate adhesin proteins of therapeutic potential. The invention particularly describes a novel method to identify adhesin proteins in different genomes of pathogens. These adhesin proteins can be used for developing preventive therapeutics.

Accordingly, a computational method for identifying adhesin and adhesin-like proteins of therapeutic potential which comprises calculation of 105 compositional properties under the five sequence attributes, namely, Amino Acid frequency, Multiplet frequency, dipeptide frequency, charge composition and hydrophobic composition; and then training Artificial Neural Network (ANN, Feed Forward Error Back Propagation)

using these properties for differentiating between adhesin and non-adhesin class of proteins. This computational method involves quantifying 105 compositional attributes of query proteins and qualifying them as adhesins or non-adhesins by a P_{ad} value (Probability of being an adhesin). The present invention is useful for identification of adhesin and adhesin-like proteins in pathogenic organisms. These newly identified adhesin and adhesin-like proteins constitute a set of candidates for development of new preventive therapeutics that can be tested in suitable experimental model systems readily. In addition, the genes encoding the candidate adhesin and adhesin-like proteins are provided.

The invention provides a set of candidate adhesin and adhesin-like proteins and their coding genes for further evaluation as preventive therapeutics. The method of invention is based on the analysis of protein sequence attributes instead of sequence patterns classified to functional domains. Present method is less dependent on sequence relationships and therefore offers the potential power of identifying adhesins from distantly related organisms. The invention provides a computational method, which involves prediction of adhesin and adhesin-like proteins using Artificial Neural Networks. The proteins termed adhesin were found to be predicted with a high probability ($P_{ad} = 0.51$) in various pathogens. Some adhesin sequences turned out to be identical or homologous to proteins that are antigenic or implicated in virulence. By this approach, proteins could be identified and short-listed for further testing in development of new vaccine formulations to eliminate diseases caused by various pathogenic organisms.

DESCRIPTION OF TABLES

Table 1: Output file format given by SPAAN.

Table 2: Organism Name, Accession number, Number of base pairs, Date of release and Total number of proteins.

Table 3. Prediction of well characterized adhesins from various bacterial pathogens using SPAAN.

Table 4. Analysis of predictions made by SPAAN on genome scans of a few selected pathogenic organisms.

Table 5: GI numbers and Gene IDs of new putative adhesins predicted by SPAAN in the genomes listed in Table 2.

Table 6: GI numbers and Gene IDs of hypothetical proteins predicted as putative adhesins by SPAAN in the genomes listed in Table 2.

Table 7: The list of 198 adhesins found in bacteria

Brief description of the accompanying drawings

5 **Figure 1** shows the Neural Network architecture

Figure 2 shows assessment of SPAAN using defined test dataset.

Figure 3 (a) shows Histogram plots of the number of proteins in the various P_{ad} value ranges are shown. **(b)** Pairwise sequence relationships among the adhesins were determined using CLUSTAL W and plotted on X-axis. Higher scores indicate similar pairs. **(c)** plot for non-adhesins. Data are plotted in the 4 quadrant format for clear inspection.

Software program was written in C Language and operated on Red Hat Linux 8.0 operating system. The computer program accepts input protein sequences in FastA format and produces a tabulated output. The output Table contains one row for each protein listing the probability outputs of each of the five modules, a weighted average probability of these five modules (P_{ad}), and the function of the protein as described in the input sequence file. This software is called SPAAN (A Software for Prediction of Adhesins and Adhesin-like proteins using Neural Networks) and a software copyright has been filed. Although this software has multiple modules, the running of these modules have been integrated and automated. The user only needs to run one command.

AAcomp.c:

Input: File containing protein sequences in the fasta format.

Output: File containing frequencies of all 20 AAs for each protein in one row.

25 **charge.c:**

Input: File containing protein sequences in the fasta format.

Output: File containing frequency of charged amino acids (R, K, E and D) and moments (up to 18th order) of the positions of charged amino acids.

hdr.c:

30 **Input:** File containing protein sequences in the fasta format.

Output: File containing frequencies of 5 groups of amino acids formed on the basis their Hydrophobicity and moments of their positions up to 5th order.

multiplets.c:

Input: File containing protein sequences in the fasta format.

5 **Output:** File containing fractions of multiplets of each of the 20 amino acids.

querydipep.c:

Input: File.1 containing protein sequences in the fasta format.

File.2 containing list of the significant dipeptides in dipeptide analysis.

10 **Output:** File containing frequencies of the dipeptides listed in the input File.2 for each protein in the input File.1.

train.c:

Input: File containing following specifications –

- 15
1. Number of input and output parameters.
 2. Number of nodes in the hidden layers.
 3. Names of the training, validate and test data files.
 4. Learning rate, coefficient of moment.
 5. Maximum number of cycles for training.

Output: Outputs are as follows.

- 20
1. Output of the trained NN for the test data set.
 2. Values of the weight connections in the trained NN.
 3. Some extra information about training.

recognize.c:

Input: File containing following specifications –

- 25
1. Number of input and output parameters.
 2. Number of nodes in the hidden layers.
 3. Names of the query input file.
 4. Name of the file containing values of the weight connections for trained NN.
 5. Name of the output file.

30 **Output:** Outputs for the query entries calculated by the trained NN.

standard.c:

Input: File containing protein sequences in fasta format.

Output: File containing protein sequences in fasta format with all the new line characters removed lying within a sequence.

5 **filter.c:**

Input: File containing protein sequences in fasta format.

Output: File containing protein sequences from the input except those which are short in length (<50 AAs) and which contain any amino acid other than the 20 known amino acids.

10 ***The five attributes:***Amino Acid frequencies

Amino acid frequency $f_i = (\text{counts of } i^{\text{th}} \text{ amino acid in the sequence}) / l$; $i = 1 \dots 20$, l is the length of the protein.

Multiplet frequency

15 Multiplets are defined as homopolymeric stretches $(X)_n$ where X is any of the 20 amino acids and n is an integer > 2 . After identifying all the multiplets, the frequencies of the amino acids in the multiplets were computed as

$$f_i(m) = (\text{counts of } i^{\text{th}} \text{ amino acid occurring as multiplet}) / l$$

Dipeptide frequencies

20 The frequency of a dipeptide (i, j) $f_{ij} = (\text{counts of } ij^{\text{th}} \text{ dipeptide}) / (\text{total dipeptide counts})$; i, j ranges from 1 to 20.

It has been found that dipeptide repeats in proteins are important for functional expression of the clumping factor present on *Staphylococcus aureus* cell surface that binds to fibrinogen (Hartford *et al* 1999). Thus we included the dipeptide frequency module. The total number of dipeptides is 400. For optimal training of Neural Network, the ratio of total number of input vectors to the total number of weight connections must be around 2 to avoid over fitting (Andrea *et al*). Therefore, we identified the dipeptides whose frequencies in the adhesin data set (469 proteins, see database construction) were significantly different from that in the non-adhesin dataset (703 proteins) using *t-test*. The frequencies of top 20 dipeptides (when arranged in the descending order of the p-values of *t-test*), were fed to the Neural Network. These dipeptides were (using single letter IUPAC-IUB code) NG, RE, TN, NT, GT, TT, DE,

ER, RR, RK, RI, AT, TS, IV, SG, GS, TG, GN, VI, AND HR. With frequency inputs for 20 dipeptides and 28 neurons in the 2nd layer, the total number of weight connections is 588, and is in keeping with the criterion of avoiding over fitting.

Charge composition

- 5 The input frequency of charged amino acids (R, K, E and D considering the ionization properties of the side chains at pH 7.2) given by $f_c = (\text{counts of charged amino acids}) / 1$. Further, information on the characteristics of the distribution of the charged amino acids in a given protein sequence was provided by computing the moments of the positions of the occurrences of the charged amino acids. Since moments characterize the patterns of distribution such as skewness and kurtosis (sharpness of the peak) we have used them to represent the distribution patterns of the charged residues in the sequence.

The general expression to compute moments of a given order; say 'i' is

$M_r = r^{\text{th}}$ order moment of the positions of charged amino acids

$$15 \quad = \sum \frac{(X_i - X_m)^r}{N}$$

Where, X_m = mean of all positions of charged amino acids

X_i = position of i^{th} charged amino acid

N = number of charged amino acids in the sequence

- The moments 2nd to 19th order were used to train the ANN constituting a total 20 inputs in addition to frequency of charged amino acids and the length of the protein. The upper limit of 19th order was set based on assessments of sensitivity and specificity on a small dataset of adhesins and non-adhesins. Moments of order greater than 19 were not useful in improvement of performance.

Hydrophobic composition

- 25 A given protein sequence was digitally transformed using the hydrophobic scores of the amino acids according to Brendel *et al.* (43). The scores for five groups of amino acids: (-8 for K, E, D, R), (-4 for S, T, N, Q), (-2 for P, H), (+1 for A, G, Y, C, W), (+2 for L, V, I, F, M).

Following inputs were given for each of the group

- 30 (a) $f_i = (\text{counts of } i^{\text{th}} \text{ group}) / (\text{total counts in the protein})$; i ranges from 1 to 5
 (b) $m_{ji} = j^{\text{th}}$ order moment of positions of amino acids in i^{th} group; j ranges from 2 to 5.

A total of 25 inputs representing the hydrophobic composition of a protein were fed to the Neural Network. The rationale for using moments was same as described in the section on charge composition inputs.

Taken together a total of 105 compositional properties of a given protein sequence were used to predict their adhesin characteristics.

The software PropSearch uses 144 compositional properties of protein sequences to detect possible structural or functional relationships between a new sequence and sequences in the database (Hobohm and Sander 1995). The approach defines protein sequence dissimilarity (or distance) as a weighted sum of differences of compositional properties such as singlet and doublet amino acid composition, molecular weight, isoelectric point (protein property search or PropSearch). Compositional properties of proteins have also been used for predicting secretory proteins in bacteria and apicoplast targeted proteins in *Plasmodium falciparum* (Zuegge, et al. 2001). The properties used here are statistical methods, principal component analysis, self-organizing maps, and supervised neural networks. In SPAAN, we have used 105 compositional properties in the five modules viz. Amino Acid frequencies, Multiplet frequencies, Dipeptide frequencies, Charge composition, Hydrophobic composition. The total of 105 properties used in SPAAN are 20 for Amino acid frequencies, 20 for Multiplets frequencies, 20 for Dipeptide frequencies (Top 20 significant dipeptides are used, based on *t-test*), 20 for Charge composition (frequency of charged amino acids (R, K, E and D) and moments of 2nd to 19th order), and 25 for Hydrophobic composition (Amino acids were classified into five groups (-8 for K, E, D, R), (-4 for S, T, N, Q), (-2 for P, H), (+1 for A, G, Y, C, W), (+2 for L, V, I, F, M). A total of 25 inputs consisted of the following: Frequency of each group, Moments of positions of amino acids in each group from 2nd to 5th order.

Neural Network

A feed forward error back propagation Neural Network was used. The program is a kind gift from Charles W. Anderson, Department of Computer Science, Colorado State University, Fort Collins, CO 80523, anderson@cs.colostate.edu

Neural Network architecture

The Neural Network used here has a multi-layer feed-forward topology. It consists of an input layer, one hidden layer and an output layer. This is a 'fully-connected' Neural Network where each neuron *i* is connected to each unit *j* of the next layer (Figure 1).

The weight of each connection is denoted by w_{ij} . The state I_i of each neuron in the input layer is assigned directly from the input data, whereas the states of hidden layer neurons are computed by the sigmoid function,

$$h_j = 1 / (1 + \exp -(w_{j0} + \sum w_{ij} I_i)),$$

5 where, w_{j0} is the bias weight

The back propagation algorithm was used to minimize the differences between the computed output and the desired output. Ten thousand cycles (epochs) of iterations are performed. Subsequently, the best epoch with minimum error was identified. At this point the network produces approximate target values for a given input in the training set.

A network was trained optimally for each attribute. Thus five networks were prepared. The schematic diagram (Figure 1) shows the procedure adopted. The number of neurons in the input layer was equal to the number of input data points for each attribute (for example 20 neurons for 20 numerical input vectors of the amino acid composition attribute). The optimal number of neurons in the hidden layer was determined through experimentation for minimizing the error at the best epoch for each network individually. An upper limit for the total number of weight connections was set to half of the total number of input vectors to avoid over fitting as suggested previously (Andrea *et al*).

20 Computer programs to compute individual compositional attributes were written in C and executed on a PC under Red Hat Linux ver 7.3 or 8.0. The network was trained on the training set, checks error and optimizes using the validate set through back propagation. The validate set was different from the training set. Since, the number of well annotated adhesins were not many, we used the 'validate set' itself as test set for preliminary evaluation of the performance and to obtain the fraction of correlation to compute the weighted average probability (P_{ad} value) described in the next section. The training set had 367 adhesins and 580 non-adhesins. The validate set had 102 adhesins and 123 non-adhesins. The adhesins were qualified with a digit '1' and the non-adhesins were qualified with a digit '0'.

30 During predictions, the network is fed with new data from the sequences that were not part of training set. Each network assigns a probability value of being an adhesin to a given sequence. The final probability is computed as described in the next section.

Probability of being an adhesin, the P_{ad} value

Query proteins are processed modularly through network trained for each attribute. Thus, five probability outputs are obtained. Final prediction was computed using the following expression which is a weighted linear sum of the probabilities from five modules:

5

$$P_{ad} = \frac{(P_A * fc_A + P_C * fc_C + P_D * fc_D + P_H * fc_H + P_M * fc_M)}{(fc_A + fc_C + fc_D + fc_H + fc_M)}$$

P_i = Probability from i module,

fc_i = fraction of correlation of i module of the trained Neural Network,

Where i = A (Amino acid frequencies), C (Charge composition), D (Dipeptide
10 frequencies), H (Hydrophobic composition), or M (Multiplet frequencies).

The fraction of correlation fc_i represents the fraction of total entries that were correctly predicted ($P_{i,adhesin} > 0.5$ and $P_{i, non-adhesin} < 0.5$) by the trained network on the test set used in preliminary evaluation (Charles Anderson).

Neural Network

15 A feed forward error back propagation Neural Network was used. The program was downloaded from the web site with permission from the author, Charles W. Anderson, Department of Computer Science, Colorado State University, Fort Collins, CO 80523, anderson@cs.colostate.edu

Statistical Analysis

20 All statistical procedures were carried out using Microsoft Excel (Microsoft Corporation Inc. USA).

Sequence analysis

Homology analysis was carried out using CLUSTAL W (Thompson *et al* 1994), BLAST (Altschul *et al* 1990), CDD (conserved domain database) search (Marchler-
25 Bauer *et al* 2002).

The whole genome sequences of microbial pathogens present new opportunities for the development of clinical applications such as diagnostics and vaccines. The present invention provides new leads for the development of candidate genes, and their encoded proteins in their functional relevance to preventive therapeutics.

30 The protein sequences of both the classes, i.e. adhesin and non-adhesin, were downloaded from the existing database (National Centre for Biotechnology Information (NCBI), USA). A total of 105 compositional properties under the five sequence

attributes namely, amino acid composition, multiplet composition, dipeptide composition, charge composition and hydrophobic composition were computed by computer programs written in C language. The attributes were computed for all the proteins in both the databases. The sequence-based attributes were then used to train Artificial Neural Network for each of the protein attributes. Adhesins were qualified by the digit '1' and non-adhesins were qualified by the digit '0'. Finally each trained Artificial Neural Network was used to identify potential adhesins which can be envisaged to be useful for the development of preventive therapeutics against pathogenic infections. Accordingly, the invention provides a computational method for identifying adhesin and adhesin-like proteins of therapeutic potential, which comprises:

1. preparing two comprehensive data-sets of adhesin and non-adhesin proteins from publicly available information on protein sequences,
2. calculating computationally the sequence based attributes of the protein sequences in the publicly available protein datasets using specially developed Software for Prediction of Adhesins and Adhesin-like proteins using Neural Networks (SPAAN),
3. training the Artificial Neural Network (ANN) for the selected attributes,
4. assigning probability value suitable for an adhesin, " P_{ad} " to the query protein and identifying adhesin like property in the query proteins with the help of trained Artificial Neural Network implemented in SPAAN,
5. validating computationally the protein sequences as therapeutic potentials by comparing with the known protein sequences that are biochemically characterized in the pathogen genome.

In an embodiment of the invention the protein sequence data may be taken from an organism, specifically but not limited to organisms such as *Escherichia coli*, *Haemophilus influenzae*, *Helicobacter pylori*, *Mycoplasma pneumoniae*, *Mycobacterium tuberculosis*, *Rickettsiae prowazekii*, *Porphyromonas gingivalis*, *Shigella flexneri*, *Streptococcus mutans*, *Streptococcus pneumoniae*, *Neisseria meningitides*, *Streptococcus pyogenes*, *Treponema pallidum*, Severe Acute Respiratory Syndrome associated coronavirus.

In another embodiment to the present invention different sequence-based attributes used for identification of proteins of therapeutic potential, comprise amino acid composition, charge composition, hydrophobicity composition, multiplets frequencies, and dipeptide frequencies.

In an embodiment, the non-homologous adhesin protein sequence may be compared with that of known sequences of therapeutic applications in the selected pathogens.

In an embodiment of the invention, the sequences of adhesin or adhesin like proteins comprise sequences of sequences IDs listed in Tables 5 and 6 identified by the method
5 of invention.

.Another embodiment of the invention the computer system comprises a central processing unit, executing SPAAN program, giving probabilities based on different attributes using Artificial Neural Network and in built other programs of assessing attributes, all stored in a memory device accessed by CPU, a display on which the
10 central processing unit displays the screens of the above mentioned programs in response to user inputs; and a user interface device.

In One embodiment of the present invention, the particulars of the organisms such as their name, strain, accession number in NCBI database and other details are given in Table 2:

15 The invention is further explained with the help of the following examples, which are given by illustration and should be construed to limit the scope of the present invention in any manner.

Example 1

Operating SPAAN:

20 The purpose of the program is to computationally calculate various sequence-based attributes of the protein sequences.

The program works as follows:

The internet downloaded FASTA format files obtained from <http://www.ncbi.nlm.nih.gov> were saved by the name <organism_name>.faa are
25 converted in the standard format by C program and passed as input to another set of C programs which computes the 5 different attributes of protein sequences (a total of 105 compositional properties in all 5 modules).

The computed properties were fed as input to the 5 different Neural Networks. Each trained network assigns a probability value of being an adhesin for a query protein. The
30 final probability (P_{ad}) was calculated as weighted average of these five individual probabilities. The weights were determined from a correlation value of correct prediction during test runs of each of the five modules.

Input/Output format:**Downloaded Files and their format:**

<organism_name>.faa: file which stores the annotation and the protein sequence.

Input file Format: FASTA

5 ">gi.vertline."<annotation>

For example,

>gi.vertline.2314605.vertline.gb.vertline.AAD08472.vertline.histidine and glutamine-rich protein

MAHHEQQQQQQANSQHSHHHHHHAHHHHYYGGEHHHHNAQQHAEQQAEQQ
10 AQQQQQQQAHQQQQQKAQQQNQQY

>gi.vertline.3261822.vertline.gn1.vertline.PID.vertline.e328405 PE_PGRS

MIGDGANGGPGQPGGPGGLLYGNGGHGGAGAAGQDRGAGNSAGLIGNGGAG
GAGGNGGIGGAGAPGGLGGDGGKGGFADEFTGGFAQGGRRGGFGGNGNTGAS
GGMGGAGGAGGAGGAGGLLIGDGGAGGAGGIGGAGGVGGGGGAGGTGGGG
15 VASAFGGGNAFGGRGGDGGDGGDGGTGGAGGARGAGGAGGAGGWLSGHSG
AHGAMGSGEGGAGGGGGARGEAGAGGGTSTGTNPGKAGAPGTQGDSGDP
GPPG

>gi.vertline.. .

Table 1: Output file format given by SPAAN

20 <organism_name>.out

SN	P _A	P _C	P _D	P _H	P _M	P _{ad} -value	Protein Name
1	0.05683	0.290803	0.441338	0.50304	0.029503	0.260485	>gi.vertline.32454344.vert line.gb.vertline.AAP82966 .1. vertline.orfla polyprotein [SARS coronavirus Hong Kong ZY-2003]
2	0.639235	0.166721	0.054583	0.935385	0.453498	0.462452	>gi.vertline.32454345.vert line.gb.vertline.AAP82967 .1. vertline.orflab polyprotein [SARS coronavirus Hong Kong ZY-2003]
3	0.65111 1	0.91150 4	0.43869 6	0.54394 4	0.92404 4	0.690247	>gi.vertline.32454346.vert line.gb.vertline.AAP82968 .1. vertline.spike glycoprotein [SARS coronavirus Hong Kong ZY-2003]

4	0.464324	0.655003	0.179503	0.008700	0.241573	0.300970	>gi.vertline.32454347.vert line.gb.vertline.AAP82969 .1. vertline.Orf3a [SARS coronavirus Hong Kong ZY-2003]
---	----------	----------	----------	----------	----------	----------	---

Where P_A , P_C , P_D , P_H , P_M are the outputs of the five Neural Networks.

Example 2 organisms and sequence numbers

Table 2: Organism Name, Accession number, Number of base pairs, Date of release
5 and Total number of proteins analyzed

Organism Name	Accession Number	Number of base pairs	Date of release	Total no. of proteins
<i>E. coli</i> O157 H7	NC_002695	5498450	7-Mar-2001	5361
<i>H. influenzae</i> Rd	NC_000907	1830138	30-Sep-1996	1709
<i>H. pylori</i> J99	NC_000921	1643831	10-Sep-2001	1491
<i>M. pneumoniae</i>	NC_000912	816394	2-Apr-2001	689
<i>M. tuberculosis</i> H37Rv	NC_000962	4411529	7-Sep-2001	3927
<i>R. prowazekii</i> strain Madrid E	NC_000963	1111523	10-Sep-2001	835
<i>P. gingivalis</i> W83	NC_002950	2343476	9-Sep-2003	1909
<i>S. flexneri</i> 2a str. 2457T	NC_004741	4599354	23- Apr-2003	4072
<i>S. mutans</i> UA159	NC_004350	2030921	25-Oct-2002	1960
<i>S. pneumoniae</i> R6	NC_003098	2038615	6-Sep-2001	2043
<i>N. meningitidis</i> serogroup A strain Z2491	NC_003116	2184406	27-Sep-2001	2065
<i>S. pyogenes</i> MGAS8232	NC_003485	1895017	Jan 31, 2002	1845
<i>T. pallidum</i> subsp. pallidum str. Nichols	NC_000919	1138011	7-Sep-2001	1036
Severe Acute Respiratory Syndrome (SARS) associated coronavirus Frankfurt 1	AY291315	29727	11-JUN-2003	14
SARS coronavirus HSR 1	AY323977	29751	15-OCT-2003	14

<i>SARS coronavirus</i> ZJ01	AY29702 8	29715	19-MAY-2003	3
<i>SARS coronavirus</i> TW1	AY29145 1	29729	14-MAY-2003	11
<i>SARS coronavirus</i> CUHK-Su10	AY28275 2	29736	07-MAY-2003	4
<i>SARS coronavirus</i> Urbani	AY27874 1	29727	12-AUG-2003	12
<i>SARS coronavirus</i>	NC_0047 18	29751	9-Sep-2003	29
<i>SARS coronavirus</i> Tor2	AY27411 9	29751	16-MAY-2003	15
<i>SARS coronavirus</i> GD01	AY27848 9	29757	18-AUG-2003	12
<i>SARS coronavirus</i> CUHK-W1	AY27855 4	29736	31-JUL-2003	11
<i>SARS coronavirus</i> BJ01	AY27848 8	29725	01-MAY-2003	11

Example 3

The multi-layered feed forward Neural Network architecture implemented in SPAAN (figure 1). A given protein sequence in FASTA format is first processed through the 5 modules A, C, D, H, and M to quantify the five types of compositional attributes. A: Amino acid composition, C: Charge composition, D: dipeptide composition of the 20 dipeptides (NG, RE, TN, NT, GT, TT, DE, ER, RR, RK, RI, AT, TS, IV, SG, GS, TG, GN, VI, HR), H: Hydrophobic composition, M: Amino acid frequencies as Multiplets. The sequence shown is part of the FimH precursor (gi 5524634) of *E. coli*. Subsequently, these numerical data are input to the input neuron layer. The directions of arrows show data flow. The number of neurons chosen in the input layer was equal to the number of the numerical input vectors of each module. The network was optimally trained through minimization of error of detection based on validate set through back propagation. The details are described in the methods. Each network module assigns a probability value of the protein being an adhesin based on the corresponding attribute. The final probability of a protein sequence being an adhesin is the P_{ad} value a weighted average of the individual probabilities and the associated fraction of correlation which is a measure of the strength of the prediction.

Example 4

Performance of SPAAN assessed using a test set of 37 adhesins and 37 non-adhesins that were not part of the training set. Matthew's correlation coefficient (Mcc, plotted on

Y-axis) for all the proteins with P_{ad} values above a given threshold (plotted on X-axis) (figure 2). The Matthew's correlation is defined as:

$$Mcc = \frac{(TP * TN) - (FP * FN)}{\sqrt{(TN + FN)(TN + FP)(TP + FN)(TP + FP)}}$$

Where TP = True Positives, TN = True Negatives, FP = False Positives, FN = False
5 Negatives.

Here TPs are adhesins, TNs are non-adhesins. In general, adhesins have high P_{ad} value, whereas non-adhesins have low P_{ad} value. Thus known adhesins with P_{ad} value above a given threshold are true positives whereas known non-adhesins with P_{ad} value below the given threshold are true negatives. The sensitivity, Sn is given by $\left(\frac{TP}{TP + FN}\right)$ and

10 specificity, Sp is given by $\left(\frac{TN}{TN + FP}\right)$. False negatives are those cases, wherein a

known adhesin had P_{ad} value lower than the chosen threshold. Similarly, a known non-adhesin with a P_{ad} value higher than the chosen threshold was taken as false positive. A theoretical polynomial curve of second order (dashed line) was fitted to the observed curve (smooth line) with a Karl-Pearson correlation coefficient $R^2 = 0.9799$. The
15 maximum point of the theoretical curve (where first derivative vanishes and second derivative is negative) was chosen as reference (vertical dotted line) to identify the maximum $Mcc = 0.94$ on the observed curve (shown by arrow). The corresponding P_{ad} value threshold was 0.51. At this P_{ad} value threshold, Sn and Sp were 0.89 and 1.0 respectively. Note that the Mcc does not drop down to the x-axis because the highest
20 P_{ad} value attained by adhesins was 0.939 in comparison to the theoretical attainable limit of 1.0.

Example 5

Assessment of SPAAN on well known adhesins from various bacterial pathogens.

Table 3. Prediction of well characterized adhesins from various bacterial pathogens
25 using SPAAN.

Species	Disease caused	Adhesin ^a	Host ligand	P_{ad} value ^b (Range)
<i>E. coli</i>	Diarrhoea	PapG (27)	α -D-gal(1-4) β -D-Gal- containing receptors	0.84-0.76
		SfaS (5)	alpha-sialyl-beta-2,3-b- galactose	0.94-0.94
		FimH (63)	D-mannosides	0.96-0.23 ^c

		Intimin (12)	tyrosine-phosphorylated form of host cell receptor Hp90	0.95-0.78
		PrsG (5)	Gal(alpha1-4)Gal	0.86-0.85
Nontypeable <i>H. influenzae</i>	Influenza	HMW1, HMW2	Human epithelial cells	0.97
		Hia (8)	human conjunctival cells	0.93-0.90
<i>H. influenzae</i>	bacterial meningitis ^d	HifE (18)	Sialylyganglioside-GM1	0.85-0.73
<i>K. pneumoniae</i>	Pneumonia	MrkD	type V collagen	0.82
<i>B. pertussis</i>	Whooping cough	FHA	Sulphated sugars on cell-surface glycoconjugates	0.85
		Pertactin	Integrins	0.43
<i>Y. enterocolitica</i>	Enterocolitis	YadA (5)	β_1 integrins	0.88-0.79
<i>S. mutans</i>	Dental Caries	SpaP (2)	Salivary glycoprotein	0.88, 0.87
		PAC	Salivary glycoprotein	0.88
<i>Streptococcus gordonii</i>	Oral cavity	SspA (2)	Salivary glycoprotein	0.85, 0.84
		CshA	Fibronectin	0.78
		CshB	Fibronectin	0.63
		ScaA	Co-aggregation	0.71
		SspB (2)	Salivary glycoprotein	0.85, 0.84
<i>Streptococcus sobrinus</i>	Tooth decay	SpaA	Salivary glycoprotein	0.89
		PAG (2)	Salivary glycoprotein	0.89, 0.73
<i>Streptococcus pyogenes</i>	Scarlet Fever	Protein F	Fibronectin	0.49
<i>Streptococcus pneumoniae</i>	Bacterial Pneumonia	PsaA (5)	Human nasopharyngeal cells	0.82-0.78
		CbpA ^e / SpsA / PbcA/ PspC	phosphorylcholine of the teichoic acid.	0.81-0.49
<i>Streptococcus parasanguis</i>	Valve endocarditis	FimA	Salivary glycoprotein fibrin	0.76
<i>Streptococcus sanguis</i>	Tooth Decay	SsaB	Salivary glycoprotein	0.71
<i>Enterococcus faecalis</i>	Empyema in patients with liver disease	EfaA	Unknown	0.83
<i>Staphylococcus aureus</i>	Food Poisoning	FnbA	Fibronectin	0.8
		FnbB (3)	Fibronectin	0.78, 0.77, 0.69
<i>Helicobacter pylori</i>	Peptic Ulcers	BabA (17)	difucosylated Lewis ^b blood group antigen	0.87-0.68

^a: The number of sequences from different strains and homologs from related species analyzed are shown in parantheses.

^b: Rounded off to the second decimal.

^c: Out of 63 FimH proteins, 54 were from *E. coli*, 6 from *Shigella flexneri*, 2 from *Salmonella enterica* and 1 was from *Salmonella typhimurium*. Except 2 FimH proteins, the rest had $P_{ad} = 0.51$. The 2 exceptions (gi numbers: 5524636, 1778448) were from *E. coli*. The gi:5524636 protein is annotated as a FimH precursor but is much shorter (129 amino acids) than other members of the family. The gi:1778448 protein is a *S. typhimurium* homolog in *E. coli*.

^d: Other ailments include pneumonia, epiglottitis, osteomyelitis, septic arthritis and sepsis in infants and older children.

^e: The adhesin CbpA is also known by alternative names SpsA, PbcA and PspC. A total of seven sequences were analyzed. Except 1 PspC sequence, the rest all had $P_{ad} = 0.51$.

Example 6

Ability of SPAAN to discriminate adhesins from non-adhesins at $P_{ad} = 0.51$ (figure 3-a).

Example 7

The non-homology character of SPAAN assesses in both adhesins and non-adhesins (figure 3b and 3c).

Figure 3 (a – c). SPAAN is non-homology based software. A total of 130 adhesins and 130 non-adhesins were analyzed to assess whether the predictive power of SPAAN could be influenced by the sequence relationships. (a) Histogram plots of the number of proteins in the various P_{ad} value ranges are shown. Shaded bars represent adhesins whereas open bars represent non-adhesins. Note the SPAAN's ability to segregate adhesins and non-adhesins into two distinct cohesive groups. (b) Pairwise sequence relationships among the adhesins were determined using CLUSTAL W and plotted on X-axis. Higher scores indicate similar pairs. The corresponding differences in P_{ad} values in the same protein pair was plotted on the Y-axis. Each point in the diagram represents a pair. Arrow points to protein pairs of the FimH family with high ΔP_{ad} values in spite of high similarity: Since one of the FimH proteins (gi: 5524636) had very low P_{ad} value all pairs with this false negative protein show high ΔP_{ad} values. The protein (gi: 5524636) is of much shorter length compared with other members of the same family. (c) plot for non-adhesins. Data are plotted in the 4 quadrant format for clear inspection. Note that among protein pairs with CLUSTAL W score < 20 the majority (82% in adhesins and 86% in non-adhesins) have $\Delta P_{ad} < 0.2$. These data support the non-homology character of SPAAN.

Example 8

Genomescan of pathogens by SPAAN identifies well known adhesins and new adhesins and adhesin-like proteins

Table 4. Analysis of predictions made by SPAAN on genome scans of a few selected

5 pathogenic organisms^a

Protein Class \ Species	<i>Escherichia coli</i> O157:H7	<i>Mycobacterium tuberculosis</i> H37Rv	SARS associated corona virus (11 strains)
Total number of proteins with $P_{ad} \geq 0.51$	575	435	5
Known adhesins	17 ^b	-	-
Putative proteins with adhesin like characteristics	92 ^c	105 ^j	-
Hypothetical proteins with adhesin-like characteristics	22 ^d	-	-
Proteins likely to be extracytoplasmic or located at surface	190 ^e	191 ^k	5 ^m
Phage proteins	30 ^f	-	-
Others	13 ^g	6 ^l	-
Hypothetical proteins	157 ^h	86 ^h	-
Wrong predictions	54 ⁱ	47 ⁱ	-

^a: SPAAN has general applicability. The three pathogens chosen here are those in which intense investigations are being conducted presently. *M. tuberculosis* is of special importance to developing countries.

10 ^b: Fimbrial adhesins, AidA-I, gamma intimin, curlin, translocated intimin receptor, putative adhesin and transport, Iha, prepilin peptidase dependent protein C.

^c: These proteins have been annotated as proteins with a putative function. These sequences were analyzed using CDD (Conserved domain database, NCBI) and BLAST searches. Adhesin like domains were found in these proteins.

15 ^d: These proteins have been annotated as 'hypothetical'. These sequences were analyzed using CDD and BLAST searches. Adhesin like domains were found in these proteins.

^e: These proteins are outer membrane, extracellular, transport, surface, exported, flagellar, periplasmic lipoprotein, and proteins annotated as 'hypothetical' but found to have similar functions listed here using BLAST and CDD searches.

^f: The phage proteins were of the following functional roles – tail fiber, head
5 decoration, DNA injection, tail, major capsid, host specificity, endolysin.

^g: Proteins predicted by SPAAN but not readily classifiable into the classes listed here have been collectively grouped as 'Others'. However, some of these proteins are known to participate in host-pathogen interactions. The annotated functional roles are typeIII secretion, antibiotic resistance, heat shock, acid shock, structural, tellurium resistance,
10 terminase, Hcp-like, Sec-independent translocase, uncharacterized nucleoprotein, HicB-like.

^h: These proteins have been annotated as hypothetical. Re-analyses of these proteins using BLAST and CDD failed identify any function for these proteins.

ⁱ: These proteins have been annotated with functional roles that are very likely to occur
15 within the cell. Hence these proteins may have remote possibility of functioning as adhesins or adhesin-like proteins. Therefore this set of proteins have been incorrectly predicted as adhesins or adhesin-like by SPAAN.

^j: These proteins are PE_PGRS, PE proteins. Several reports (for example Brennan *et al.*) indicate that PE_PGRS proteins may be localized to cell surface and aid in host-
20 pathogen interaction.

^k: Lipoproteins (lpp, lpq, lpr), PPE, outer membrane, surface, transport, secreted, periplasmic, extracellular, ESAT-6, peptidoglycan binding, exported, mpt (with extracellular domains), and proteins annotated as 'hypothetical' but found to have similar functions listed here using BLAST and CDD searches.

^l: These proteins were of the following functions - glutaredoxin-like thioltransferase, putative involvement in molybdate uptake, ATP synthase chain, sulphotransferases, S.erythraea rhodanese-like protein M29612|SERCYSA_5, unknown function.
25

^m: These proteins were the spike glycoprotein with antigenic properties, and nsp2, nsp5, nsp6 and nsp7.

Table 5: New putative adhesins predicted by SPAAN in the genomes listed in table 2 –

(Total number = 279)

Protein GI Number	Gene ID	Protein name
<i>Escherichia coli</i> O157:H7		
13360742	912619	hemagglutinin/hemolysin-related protein
13362986	914770	putative ATP-binding component of a transport system
13361114	913228	putative tail fiber protein
13364757	913676	minor fimbrial subunit/D-mannose specific adhesin
13362687	915687	putative fimbrial-like protein
13360856	912599	AidA-I adhesin-like protein
13364140	915374	putative fimbrial protein
13359793	914435	putative invasin
13364768	913650	putative invasin
13364034	915471	Gamma intimin
13362703	915668	putative DNA transfer protein precursor
13364141	915376	putative fimbrial protein
13359819	914463	AidA-I adhesin-like protein
13360480	917768	putative fimbrial-like protein
13362692	915681	putative fimbrial-like protein
13362585	916824	putative ATP-binding component of a transport system
13359881	914526	putative flagellin structural protein
13361579	917311	putative type 1 fimbrial protein precursor
13360880	913991	curlin major subunit CsgA
13364036	915465	translocated intimin receptor Tir
13360740	912615	putative major pilin protein
13361582	917317	putative ATP-binding component of a transport system and adhesin protein
13364754	913683	export and assembly outer membrane protein of type 1 fimbriae
13360484	917767	homolog of Salmonella FimH protein

13364751	913688	major type 1 subunit fimbrin
13359597	913742	putative fimbrial protein
13362550	916787	putative ATP-binding component of a transport system
13359595	913739	putative fimbrial protein
13359599	913748	probable outer membrane porin protein involved in fimbrial assembly
13363900	915704	putative fimbrial protein precursor
13361575	917307	putative fimbrial-like protein
13364756	913678	fimbrial morphology
13363496	916142	truncated putative fimbrial protein
13359601	913761	putative fimbrial-like protein
13364145	915368	putative type 1 fimbrial protein
13363902	915708	putative outer membrane usher protein precursor
13361576	917309	putative outer membrane protein
13361013	913353	putative major tail subunit
13364755	913682	fimbrial morphology
13360738	912793	putative outer membrane usher protein
13363928	915608	alpha-amylase
13363495	916144	putative outer membrane protein
13362383	916617	putative type-1 fimbrial protein
13364373	914972	outer membrane vitamin B12 receptor protein BtuB
13360879	912479	minor curlin subunit precursor CsgB
13360739	912756	putative chaperone protein
13361574	917314	putative fimbrial-like protein
13361127	913212	outer membrane protease precursor
13363210	916442	putative lipoprotein
13361104	913238	major tail protein
13361709	917446	putative major tail subunit
13359725	914366	outer membrane pore protein PhoE
13360875	913765	curli production assembly/transport component CsgF
13362170	913927	putative outer membrane protein
13361473	917203	putative BigB-like protein

13364025	915286	EspF protein
13360081	916982	outer membrane receptor for ferric enterobactin (enterochelin) and colicins B and D
13362977	914779	hypothetical lipoprotein
13360351	917632	outer membrane protein X
13360696	914208	putative outer membrane precursor
13361456	917206	putative outer membrane protein
13361626	917374	putative outer host membrane protein precursor
13361698	917449	putative outer membrane protein
13362186	913421	putative outer membrane protein precursor
13362697	915676	long-chain fatty acid transport protein FadL
13360918	914188	flagellar hook protein FlgE
13360737	912506	putative outer membrane protein
13360342	917629	putative outer membrane receptor for iron transport
13363396	916248	outer membrane channel TolC
13361958	912705	putative scaffolding protein in the formation of a murein-synthesizing holoenzyme
13359921	914566	nucleoside-specific channel-forming protein TSX
13360944	913890	outer membrane receptor for ferric iron uptake
13359998	914644	putative outer membrane transport protein
13363390	916251	putative ferrichrome iron receptor precursor
13364227	915153	outer membrane phospholipase A
13361982	912846	putative outer membrane protein
13360129	917032	a minor lipoprotein
13361817	912692	putative outer membrane protein
13360233	917507	membrane spanning protein TolA
13362837	915218	putative outer membrane lipoprotein
13362328	912985	putative colanic acid biosynthesis glycosyl transferase

Haemophilus influenzae Rd

16272254	949521	prepilin peptidase-dependent protein D
16272928	950762	immunoglobulin A1 protease
16272129	951072	lipoprotein

16273251	950616	hemoglobin-binding protein
30995429	950130	opacity protein
16272854	949634	protective surface antigen D15
16272283	950648	opacity associated protein
16272604	949701	hemoglobin-binding protein
<i>Helicobacter pylori</i> J99		
4155101	889167	putative vacuolating cytotoxin (VacA) paralog
4154798	890022	putative vacuolating cytotoxin (VacA) paralog
4155426	890036	putative vacuolating cytotoxin (VacA) paralog
4155390	890075	vacuolating cytotoxin
4155400	890058	outer membrane protein - adhesin
4155681	889718	putative Outer membrane protein
4155420	890042	Outer membrane protein/porin
4155775	889799	outer membrane protein - adhesin
4155419	890044	Outer membrane protein/porin
4154526	889066	putative Outer membrane protein
4154724	889419	putative Outer membrane protein
4155862	890404	putative Outer membrane protein
4156048	889958	putative IRON(III) DICITRATE TRANSPORT PROTEIN
4154510	889297	putative Outer membrane protein
4155432	889515	putative outer membrane protein
4155623	889671	putative Outer membrane protein
4155700	889739	putative Outer membrane function
4154740	889426	Outer membrane protein/porin
4155692	889743	putative Outer membrane protein
4155594	889648	putative outer membrane protein
4155680	889719	putative Outer membrane protein
4155217	890243	putative Outer membrane protein
4155958	889905	putative Outer membrane protein
4155201	890259	putative Outer membrane protein
4155013	889232	cag island protein
4154974	889032	putative Outer membrane protein

4155214	890244	putative Outer membrane protein
4154973	889042	Outer membrane protein
4155344	890115	putative Outer membrane protein
4155099	889160	FLAGELLIN A
4155023	888978	cag island protein
4155035	889201	cag island protein, CYTOTOXICITY ASSOCIATED IMMUNODOMINANT ANTIGEN
4155289	890164	NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ PRECURSOR

Mycoplasma pneumoniae

13507881	877207	involved in cytodherence
13507880	877268	ADP1_MYCPN adhesin P1
13508228	877211	species specific lipoprotein
13508181	877124	species specific lipoprotein
13508179	877071	Mollicute specific lipoprotein, MG307 homolog, from M. genitalium
13508178	877118	Mollicute specific lipoprotein, MG307 homolog, from M. genitalium,
13508176	876797	Mollicute specific lipoprotein, MG307 homolog, from M. genitalium
13508175	876848	Mollicute specific lipoprotein, MG307 homolog, from M. genitalium
13508106	876953	involved in cytodherence
13508350	877112	similar to phosphate binding protein Pst

Mycobacterium tuberculosis H37 Rv

15607496	886491	PPE
15607445	886592	PPE
15610644	888270	PE_PGRS
15608588	886605	PE_PGRS
15609627	887941	PE_PGRS
15610643	888256	PE_PGRS
15607718	887725	PE_PGRS

15609054	885362	PPE
15610486	888113	PPE
15610483	888120	PPE
15610479	888033	PPE
15609771	888573	PE_PGRS
15610648	888306	PE_PGRS
15610481	888114	PE_PGRS
15608117	885264	PE_PGRS
15607973	885391	PE_PGRS
15608231	885258	PE_PGRS
15608906	885429	PE_PGRS
15608891	885544	PPE
15609990	888171	PE_PGRS
15609055	885506	PPE
15608227	887094	PE_PGRS
15610524	888151	PE_PGRS
15609490	886003	PPE
15607886	888664	PE_PGRS
15609624	887909	PE_PGRS
15607420	886621	PE_PGRS
15608897	885325	PE_PGRS(wag22)
15608590	886595	PE_PGRS
15609728	887992	PE_PGRS
15608012	885742	PE_PGRS
15608534	886745	PE_PGRS
15608940	885730	PE_PGRS
15607887	888662	PE_PGRS
15609235	888312	PE_PGRS
15610694	887822	PPE
15609533	885517	PE_PGRS
15610480		PE_PGRS

Rickettsia prowazekii strain Madrid E

15604316	883411	CELL SURFACE ANTIGEN (sca3)
----------	--------	-----------------------------

15604546	883694	CELL SURFACE ANTIGEN (sca5)
----------	--------	-----------------------------

Porphyromonas gingivalis W83

34541453	2551934	hemagglutinin protein HagA
----------	---------	----------------------------

34540040	2551409	lipoprotein, putative
----------	---------	-----------------------

34540364	2552375	extracellular protease, putative
----------	---------	----------------------------------

34541613	2552074	hemagglutinin protein HagE
----------	---------	----------------------------

34540183	2551891	internalin-related protein
----------	---------	----------------------------

Shigella flexneri 2a str. 2457T

30065424	1080663	minor fimbrial subunit, D-mannose specific adhesin
----------	---------	--

30062726	1077662	putative adhesion and penetration protein
----------	---------	---

30063758	1078834	putative fimbrial-like protein
----------	---------	--------------------------------

30065431	1080671	major type 1 subunit fimbrin (pilin)
----------	---------	--------------------------------------

30063366	1078379	flagellar protein FlhD
----------	---------	------------------------

30064308	1079668	outer membrane fluffing protein
----------	---------	---------------------------------

30062613	1077555	flagellar hook protein FlgE
----------	---------	-----------------------------

30061954	1076843	conserved hypothetical lipoprotein
----------	---------	------------------------------------

30065173	1080393	putative lipase
----------	---------	-----------------

30065425	1080664	minor fimbrial subunit, precursor polypeptide
----------	---------	---

30064485	1079637	putative fimbrial protein
----------	---------	---------------------------

30062615	1077558	flagellar basal body L-ring protein FlgH
----------	---------	--

30064307	1079452	outer membrane fluffing protein
----------	---------	---------------------------------

30065601	1080859	putative glycoprotein/receptor
----------	---------	--------------------------------

30062118	1077025	putative fimbrial-like protein
----------	---------	--------------------------------

30064099	1079223	lipoprotein
----------	---------	-------------

30062616	1077559	flagellar basal body P-ring protein FlgI
----------	---------	--

30063546	1078596	putative fimbrial-like protein
----------	---------	--------------------------------

30062940	1077910	putative outer membrane protein
----------	---------	---------------------------------

30065426	1080665	minor fimbrial subunit, precursor polypeptide
----------	---------	---

30062779	1077721	putative outer membrane protein
----------	---------	---------------------------------

30064194	1079329	putative lipoprotein
----------	---------	----------------------

30063365	1078378	flagellin
----------	---------	-----------

30062298	1077222	outer membrane protein X
30064968	1080175	putative major fimbrial subunit
30061858	1076740	outer membrane pore protein E (E,Ic,NmpAB)
30062178	1080410	minor lipoprotein
30062479	1077412	putative fimbrial-like protein
30062565	1077506	minor curlin subunit precursor
30063880	1078972	putative outer membrane lipoprotein
30064531	1079686	cytoplasmic membrane protein
30065033	1080243	putative receptor protein
<i>Streptococcus mutans</i> UA159		
24378550	1029610	putative secreted antigen GbpB/SagA; putative peptidoglycan hydrolase
24379087	1028055	cell surface antigen SpaP
24380463	1029310	putative membrane protein
24379075	1028046	penicillin-binding protein 2b
24378955	1027967	penicillin-binding protein 1a; membrane carboxypeptidase
24379801	1028662	glucan-binding protein C, GbpC
24379528	1029536	hypothetical protein; possible cell wall protein, WapE
24379231	1028158	putative glucan-binding protein D; BglB-like protein
24380488	1029325	conserved hypothetical protein; possible transmembrane protein
24380291	1029139	putative amino acid binding protein
24379342	1028247	putative penicillin-binding protein, class C; fnt-like protein
24380047	1028904	putative ABC transporter, branched chain amino acid-binding protein
24378698	1029755	putative ABC transporter, metal binding lipoprotein; surface adhesin precursor; saliva-binding protein; lipoprotein receptor LraI (LraI family)
24378708	1029768	putative transfer protein
24379427	1028331	cell wall-associated protein precursor WapA
24379272	1028196	putative amino acid transporter, amino acid-binding protein
24379641	1028511	putative ABC transporter, amino acid binding protein

Streptococcus pneumoniae R6

15902395	934801	Choline-binding protein
15902381	934810	Choline-binding protein F
15902165	932894	Surface protein pspA precursor
15904047	934859	Choline binding protein D
15904036	933487	Choline binding protein A
15903986	933069	Choline-binding protein
15903796	933669	Autolysin (N-acetylmuramoyl-L-alanine amidase)

Neisseria meningitidis Z2491

15794121	907145	putative membrane protein
15794144	907168	putative surface fibril protein
15793284	906275	truncated pilin
15793460	906456	IgA-specific serine endopeptidase
15793282	906273	fimbrial protein precursor (pilin)
15793337	906332	adhesin
15793253	906243	putative lipoprotein
15794356	907848	putative lipoprotein
15793684	906699	putative membrane protein
15793290	906281	truncated pilin
15793283	906274	truncated pilin
15793475	906471	haemoglobin-haptoglobin-utilization protein
15793406	906401	porin, major outer membrane protein P.I
15794985	907333	adhesin MafA2
15794344	907836	putative lipoprotein
15794622	908118	hypothetical outer membrane protein
15793599	906604	pilus-associated protein
15793763	906779	putative periplasmic binding protein

Streptococcus pyogenes MGAS8232

19745214	995235	putative secreted protein
19746570	994224	putative penicillin-binding protein 1a
19745593	994771	putative 42 kDa protein
19745813	993958	putative adhesion protein

19745225	994839	putative choline binding protein
19745828	995250	streptolysin S associated protein
19746229	995021	putative minor tail protein
19746909	994105	putative laminin adhesion
19745560	995061	putative cell envelope proteinase

Treponema pallidum subsp. *pallidum* str. Nichols

15639714	2611034	flagellar hook protein (flgE)
15639609	2611657	tpr protein J (tprJ)
15639111	2610909	tpr protein C (tprC)
15639125	2610968	tpr protein D (tprD)

SARS coronavirus

31581505		spike protein S [SARS coronavirus Frankfurt 1]
32187357		spike protein S [SARS coronavirus HSR 1]
32187342		spike glycoprotein [SARS coronavirus ZJ01]
30698329		putative spike glycoprotein S [SARS coronavirus TW1]
30421454		putative spike glycoprotein [SARS coronavirus CUHK-Su10]
30027620		S protein [SARS coronavirus Urbani]
29836496	1489668	E2 glycoprotein precursor; putative spike glycoprotein [SARS coronavirus]
30795145		spike glycoprotein [SARS coronavirus Tor2]
31416295		spike glycoprotein S [SARS coronavirus GD01]
30023954		putative E2 glycoprotein precursor [SARS coronavirus CUHK-W1]
30275669		spike glycoprotein S [SARS coronavirus BJ01]
29837498		3C-like proteinase nsp5-pp1a/pp1ab (3CL-PRO) [SARS coronavirus]
29837501		putative nsp8-pp1a/pp1ab [SARS coronavirus]
29837503		putative nsp10-pp1a/pp1ab; formerly known as growth-factor-like protein [SARS coronavirus]
29837502		putative nsp9-pp1a/pp1ab [SARS coronavirus]

Table 6: Hypothetical proteins predicted as putative adhesins by SPAAN in the genomes listed in table 2 –

(Total number of proteins = 105)

Protein GI Gene ID
number

Escherichia coli O157:H7

13363955	915578
13360000	914929
13362244	912369
13359999	914888
13361583	917316
13361172	913156
13361131	913207
13359780	914422
13360571	912499
13362197	912893
13362260	912399
13360947	913505
13361464	917196
13361635	917367
13362421	916655
13361463	917195

Haemophilus influenzae Rd

16272115	951058
30995442	950581

Helicobacter pylori J99

4155526	889586
4155712	889748
4155632	889684
4156035	889468
4155499	

Mycoplasma pneumoniae

13507870	877230
13508239	877245
13508109	876868
13508025	877084
13507838	876784
13507883	877183
13507871	877239
13507944	877056
13508241	876750
13507942	877055
13507840	877387
13507867	877242
13508201	877044
13507941	876985
13508114	877397

Mycobacterium tuberculosis H37Rv

15611014	886198
15610173	887320
15609513	885515
15608094	885411
15610958	886155
15607528	886436
15607678	887473
15609587	885760
15610708	887227
15609526	885246
15611033	886225
15609028	885094
15607730	887771
15609121	885813
15608255	885951

15608409 887039

15609124 885815

15607734 887797

Rickettsia prowazekii strain Madrid E

15604649 883964

15604322 883472

15604659 883996

15604417 883217

Porphyromonas gingivalis W83

34540233 2551594

Shigella flexneri 2a str. 2457T

30062687 1077638

30062956 1080449

30063681 1078754

30065435 1080675

30063891 1078983

30063211 1078195

30065233 1080463

30064387 1079531

30062638 1077590

30065236 1080466

30061839 1076721

Streptococcus mutans UA159

24378864 1029452

24380475 1029319

24380237 1029088

24379203 1028139

24380480 1029320

24379275 1029489

24379291 1028216

24379295 1028215

24379804 1028663

24379162	1029417
24378987	1029363
24379179	1028118
24379166	1028107
24378827	1029444
24380216	1029067

Streptococcus pneumoniae R6

15902140	932867
15903446	934616
15903916	934001
15903848	933609
15902832	934332
15902372	934804
15902152	932889

Neisseria meningitidis Z2491

15793668	906680
15794714	907603

Streptococcus pyogenes MGAS8232

19747011	993608
19747024	994165
19747012	994373
19746396	995057
19746651	993824
19745883	995045
19745912	994077

Treponema pallidum subsp. *pallidum* str. Nichols

15639844	2611061
15639720	2611059

Table 7: The list of 198 adhesins found in bacteria
PapG (*E. coli*)

12837502

7407201

7407207
7407205
147096
4240529
7407203
42308
7443327
78746
18265934
26111419
26250987
26109826
26249418
13506767
42301
78745
129622
147092
13506906
7407209
147080
281926
7407199
147100
78744

SfaS (E.coli)

477910
264035
42959
134449
96425

FimH (E.coli)

26251208
26111640
5524634
29422425
5524630
29422435
29422415
10946257
29422419
11120564
29422457
11120562
29422459
5524632
29422455
29422453
29422451
29422449
29422447
29422445
29422443
29422437
29422433
29422431
29422429
29422427
29422423
29422421
29422417
729494
1361011
1790775

3599571
29422441
12620398
29422439
5524628
1787779
1742472
1742463
15801636
25321294
12515169
11120566
24051859
24112911
13360484
15800801
15830279
25392018
25500156
12514120
1787173
16128908
16501811
16759519
24051219
24112354
30040724
30062478
6650093
5524636
1778448

Intimin (E.coli)

	17384659
	4388530
	1389879
	15723931
	4323336
	4323338
	4323340
	4323342
	4323344
	4323346
	4323348
	4689314
PrsG (E.coli)	
	42523
	42529
	7443328
	7443329
	1172645
HMW1 (Nontypeable <i>H. influenzae</i>)	
	282097
HMW2 (Nontypeable <i>H. influenzae</i>)	
	5929966
Hia (Nontypeable <i>H. influenzae</i>)	
	25359682
	25359489
	25359709
	25359628
	25359414
	25359389
	21536216
	25359445
HifE (<i>H. influenzae</i>)	

	<i>13506868</i>
	13506870
	13506872
	13506874
	13506876
	3688787
	3688790
	3688793
	2126301
	1170264
	1170265
	533127
	535169
	3025668
	3025670
	3025672
	3025674
	642038
MrkD (<i>K. pneumoniae</i>)	
	<i>127307</i>
FHA (<i>B. pertussis</i>)	
	<i>17154501</i>
Pertactin (<i>B. pertussis</i>)	
	<i>33571840</i>
YadA (<i>Y. enterocolitica</i>)	
	<i>10955604</i>
	4324391
	28372996
	23630568
	32470319
SpaP (<i>S. mutans</i>)	
	<i>26007028</i>

	47267
PAC (S. mutans)	
	129552
SspA (Streptococcus gordonii)	
	25990270
	1100971
CshA (Streptococcus gordonii)	
	457707
CshB (Streptococcus gordonii)	
	18389220
ScaA (Streptococcus gordonii)	
	310633
SspB (Streptococcus gordonii)	
	25055226
	3220006
SpaA (Streptococcus sobrinus)	
	546643
PAG (Streptococcus sobrinus)	
	217036
	47561
Protein F (Streptococcus pyogenes)	
	19224134
PsaA (Streptococcus pneumoniae)	
	18252614
	7920456
	7920458
	7920460
	7920462
CbpA ^e / SpsA / PbcA/ PspC (Streptococcus pneumoniae)	
	14718654
	2425109

	2576331
	2576333
	3153898
	9845483
	19548141
FimA (Streptococcus parasanguis)	
	97883
SsaB (Streptococcus sanguis)	
	97882
EfaA (Enterococcus faecalis)	
	493017
FnbA (Staphylococcus aureus)	
	120457
FnbB (Staphylococcus aureus)	
	581562
	21205592
	13702452
BabA (Helicobacter pylori)	
	13309962
	13309964
	13309966
	13309968
	13309970
	13309972
	13309974
	13309976
	13309978
	13309980
	13309982
	13309984
	13309986
	13309988

13309990

13309992

13309994

Advantages:

1. The method helps in discovering putative adhesins, which are of great importance in drug discoveries and preventive therapeutics.
2. The method is useful in predicting the adhesive nature of even unique proteins,
5 because it is independent of the homology of the query proteins with other proteins.
3. This method is easy to use. For calculating the output, only the amino acid sequence is required as input. No other information is required to get the information about its adhesive nature.

REFERENCES MAY BE MADE TO

1. Andrea, T.A., Kalayeh, H.(1991) Applications of neural networks in quantitative structure-activity relationships of dihydrofolate reductase inhibitors. J. Med. Chem. 34, 2824-2836.
- 5 2. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.(1990) Basic local alignment search tool. J Mol Biol. 215(3), 403-410.
3. Bassinet L, Gueirard P, Maitre B, Housset B, Gounon P, Guiso N.(2000) Role of adhesins and toxins in invasion of human tracheal epithelial cells by Bordetella pertussis. Infect Immun. 68(4), 1934-1941.
- 10 4. Bock, K., *et al.*(1985). Specificity of binding of a strain of uropathogenic Escherichia coli to Gal alpha 1----4Gal-containing glycosphingolipids. J. Biol. Chem. 260, 8545-8551.
5. Brendel, V., Bucher, P., Nourbakhsh, I.R., Edwin Blaisdell, B., and Karlin, S.(1992) Methods and algorithms for statistical analysis of protein sequences.
- 15 Proc. Natl. Acad. Sci. USA 89, 2002-2006.
6. Brennan, M.J., Delogu, G., Chen, Y., Bardarov, S., Kriakov, J., Alavi, M., Jacobs, W.R., (2001).
7. Evidence that Mycobacterial PE_PGRS proteins are cell surface constituents that influence interactions with other cells. Infect. Immun, 69, 7326-7333.
- 20 8. De BK, Woolfitt AR, Barr JR, Daneshvar MI, Sampson JS, Ades EW, Carlone GM. (2003) Analysis of recombinant acylated pneumococcal surface adhesin A of Streptococcus pneumoniae by mass spectrometry. Arch Biochem Biophys. 15, 419(2), 147-157.
9. Eglund PG, Du LD, Kolenbrander PE (2001) Identification of independent
- 25 Streptococcus gordonii SspA and SspB functions in coaggregation with Actinomyces naeslundii. Infect Immun. 69(12), 7512-7516
10. Finlay, B.B. and Falkow, S.(1997) Common themes in microbial pathogenicity revisited. Microbiol. Mol. Biol. Rev. 61, 136-169,
11. Fraser, C.M., Eisen, J., Fleischmann, R.D., Ketchum, K.A., Peterson, S.(2000)
- 30 Comparative genomics and understanding of microbial biology. Emerg. Infect. Dis. 6,505-6512
12. Halperin, S. A., Scheifele, D., Mills, E., Guasparini, R., Humphreys, G., Barreto, L., Smith, B.(2003) Nature, evolution, and appraisal of adverse events

- and antibody response associated with the fifth consecutive dose of a five-component acellular pertussis-based combination vaccine. *Vaccine* 21, 2298–2306.
13. Hartford O, McDevitt D, Foster TJ.(1999) Matrix-binding proteins of *Staphylococcus aureus*: functional analysis of mutant and hybrid molecules. *Microbiology*. 145 (Pt 9), 2497-2505.
 14. Hobohm, U. and Sander, C.(1995) A sequence property approach to searching protein databases. *J. Mol. Biol.* 251, 390-399.
 15. Ide T, Michgehl S, Knappstein S, Heussipp G, Schmidt MA.(2003) Differential modulation by Ca²⁺ of type III secretion of diffusely adhering enteropathogenic *Escherichia coli*. *Infect Immun.* 71(4), 1725-1732.
 16. Langermann S *et al.*(2000) Vaccination with FimH adhesin protects cynomolgus monkeys from colonization and infection by uropathogenic *Escherichia coli*. *J. Infect. Dis.* 181, 774-778.
 17. Lowe A.M., Lambert, P.A., Smith, A.W.(1995) Cloning of an *Enterococcus faecalis* endocarditis antigen: homology with adhesins from some oral streptococci. *Infect Immun.* 63, 703-706.
 18. Maurer, L., Orndorff, P.(1987). Identification and characterization of genes determining receptor binding and pilus length of *Escherichia coli* type 1 pili. *J. Bacteriol.* 169, 640-645
 19. Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY, Bryant SH. (2002) CDD: a database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Res.* 1, 30(1), 281-283.
 20. Neubauer H, Hensel A, Aleksic S, Meyer H.(2000) Evaluation of a *Yersinia* adhesion gene (*yadA*) specific PCR for the identification of enteropathogenic *Yersinia enterocolitica*. *Int J Food Microbiol.* 15, 57(3), 225-227.
 21. Nishikawa, K., Kubota, Y. and Ooi, T.(1983) Classification of proteins into groups based on amino acid composition and other characters. II. grouping into four types. *J. Biochem.* 94, 997-1007.
 22. Peregrin-Alvarez, J.M., Tsoka, S., Ouzounis, C.A.(2003) The phylogenetic extent of metabolic enzymes and pathways. *Genome Res.* 13, 422-427.

23. Prinz, C., Hafsi, N. Volland, P.(2003) *Helicobacter pylori* virulence factors and the host immune response: implications for therapeutic vaccination. *Trends in Microbiol.* 11, 134-138.
24. Rapola, S., Jääntti, V., Eerola, M., Helena Mäkelä, P., Käyhty, H., Kilpi, T.(2003) Anti-PsaA and the risk of pneumococcal AOM and carriage. *Vaccine* 21, 3608–3613.
25. Rison. S.C., Teichmann, S.A., Thornton, J.M.(2002) Homology, pathway distance and chromosomal localization of the small molecule metabolism enzymes in *Escherichia coli*. *J. Mol. Biol.* 318, 911-932
- 10 26. Sperandio V, Bailey C, Giron JA, DiRita VJ, Silveira WD, Vettore AL, Kaper JB.(1996) Cloning and characterization of the gene encoding the OmpU outer membrane protein of *Vibrio cholerae*. *Infect Immun.* 64(12), 5406-5409.
27. St Geme JW 3rd, Cutter D.(2000) The *Haemophilus influenzae* Hia adhesin is an autotransporter protein that remains uncleaved at the C terminus and fully cell associated. *J Bacteriol.* 182(21), 6005-6013.
- 15 28. Thompson, J.D., Higgins, D.G., Gibson, T.J.(1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673-4680
- 20 29. Van Schilfgaarde M, van Ulsen P, Eijk P, Brand M, Stam M, Kouame J, van Alphen L, Dankert J.(2000) Characterization of adherence of nontypeable *Haemophilus influenzae* to human epithelial cells. *Infect Immun.* 68(8), 4658-4665.
30. Witzemann, T.M., Adamou, J.E., Langermann, S.(1999). Adhesins as targets for vaccine development. *Emerg. Infect. Dis.* 5, 395-403,
- 25 31. Wolf, Y.I., Rogozin, I.B., Kondrashov, A.S., and Koonin, E.V.(2001) Genome alignment, evolution of prokaryotic genome organization and prediction of gene function using genomic context. *Genome Res.* 11, 356-372
32. Yu J, Leung WK, Go MY, Chan MC, To KF, Ng EK, Chan FK, Ling TK, Chung SC, Sung JJ.(2002) Relationship between *Helicobacter pylori* babA2 status with gastric epithelial cell turnover and premalignant gastric lesions. *Gut.* 51(4), 480-484.
- 30

33. Zuegge, J., Ralph, S., Schmuker, M., McFadden, G.I., Schneider, G.(2001)
Deciphering apicoplast targeting signals--feature extraction from nuclear-
encoded precursors of Plasmodium falciparum apicoplast proteins. Gene 280,
19-26.

Claims

1. A computational method for identifying adhesin and adhesin-like proteins, said method comprising steps of:
 - a. computing the sequence-based attributes of protein sequences using five attribute modules of a neural network software, wherein the attributes are, (i) amino acid frequencies, (ii) multiplet frequency, (iii) dipeptide frequencies, (iv) charge composition, and (v) hydrophobic composition,
 - b. training a artificial neural Network (ANN) for each of the computed five attributes, and
 - c. identifying the adhesin and adhesin-like proteins having probability of being an adhesin (P_{ad}) as ≥ 0.51 .
2. A method as claimed in claim 1, wherein the protein sequences are obtained from pathogens, eukaryotes, and multicellular organisms.
3. A method as claimed in claim 1, wherein the protein sequences are obtained from the pathogens selected from a group of organisms comprising *Escherichia coli*, *Haemophilus influenzae*, *Helicobacter pylori*, *Mycoplasma pneumoniae*, *Mycobacterium tuberculosis*, *Rickettsiae prowazekii*, *Porphyromonas gingivalis*, *Shigella flexneri*, *Streptococcus mutans*, *Streptococcus pneumoniae*, *Neisseria meningitides*, *Streptococcus pyogenes*, *Treponema pallidum* and Severe Acute Respiratory Syndrome associated human coronavirus (SARS).
4. A method as claimed in claim 1, wherein the method is a non-homology method.
5. A method as claimed in claim 1, wherein the method uses 105 compositional properties of the sequences.
6. A method as claimed in claim 1, wherein the method shows sensitivity of at least 90%.
7. A method as claimed in claim 1, wherein the method shows specificity of 100%.
8. A method as claimed in claim 1, wherein the method helps identifies adhesins from distantly related organisms.
9. A method as claimed in claim 1, wherein the neural network has multi-layer feed forward topology, consisting of an input layer, one hidden layer, and an output layer.

10. A method as claimed in claim 9, wherein the number of neurons in the input layer are equal to the number of input data points for each attribute.
11. A method as claimed in claim 1, wherein the "P_{ad}" is a weighted linear sum of the probabilities from five computed attributes.
- 5 12. A method as claimed in claim 1, wherein each trained network assigns a probability value of being an adhesin for the protein sequence.
13. A computer system for performing the method of claim 1, said system comprising a central processing unit, executing SPAAN program, giving probabilities based on different attributes using Artificial Neural Network and in
10 built other programs of assessing attributes, all stored in a memory device accessed by CPU, a display on which the central processing unit displays the screens of the above mentioned programs in response to user inputs; and a user interface device.
14. A set of 274 annotated genes encoding adhesin and adhesin-like proteins,
15 having SEQ ID Nos. 385 to 658.
15. A set of 105 hypothetical genes encoding adhesin and adhesin-like proteins, having SEQ ID Nos. 659 to 763.
16. A set of 279 annotated adhesin and adhesin-like proteins of SEQ ID Nos. 1 to 279.
- 20 17. A set of 105 hypothetical adhesin and adhesin-like proteins of SEQ ID Nos. 280 to 384.
18. A fully connected multilayer feed forward Artificial Neural Network based on the computational method as claimed in claim 1, comprising of an input layer, a hidden layer and an output layer which are connected in the said sequence,
25 wherein each neuron is a binary digit number and is connected to each neuron of the subsequent layer for identifying adhesin or adhesin like proteins, wherein the program steps comprise:-
 - [a] feeding a protein sequence in FASTA format;
 - [b] processing the sequence obtained in step [a] through the 5 modules
30 named A, C, D, H and M, wherein attribute A represents an amino acid composition, attribute C represents a charge composition, attribute D represents a dipeptide composition of the 20 dipeptides [NG, RE, TN, NT, GT, TT, DE, ER, RR, RK, RI, AT, TS, IV, SG, GS, TG, GN, VI

and HR], attribute H represents a hydrophobic composition and attribute M represents amino acid frequencies in multiplets to quantify 5 types of compositional attributes of the said protein sequence to obtain numerical input vectors respectively for each of the said attributes wherein the sum of numerical input vectors is 105;

[c] processing of the numerical input vectors obtained in step [b] by the input neuron layer to obtain signals, wherein the number of neurons is equal to the number of numerical input vectors for each attribute;

[d] processing of signals obtained from step [c] by the hidden layer to obtain synaptic weighted signals, wherein the optimal number of neurons in the hidden layer was determined through experimentation for minimizing the error at the best epoch for each network individually;

[e] delivering synaptic weighted signals obtained from step [d] to the output layer for assigning of a probability value for each protein sequence fed in step [a] as being an adhesin by each network module; and

[f] using the individual probabilities obtained from step [e] for computing the final probability of a protein sequence being an adhesin denoted by the P_{ad} value, which is a weighted average of the individual probabilities obtained from step [e] and the associated fraction of correlation which is a measure of the strength of the prediction.

19. A network as claimed in claim 18, wherein the input neuron layer consists of a total of 105 neurons corresponding to 105 compositional properties.

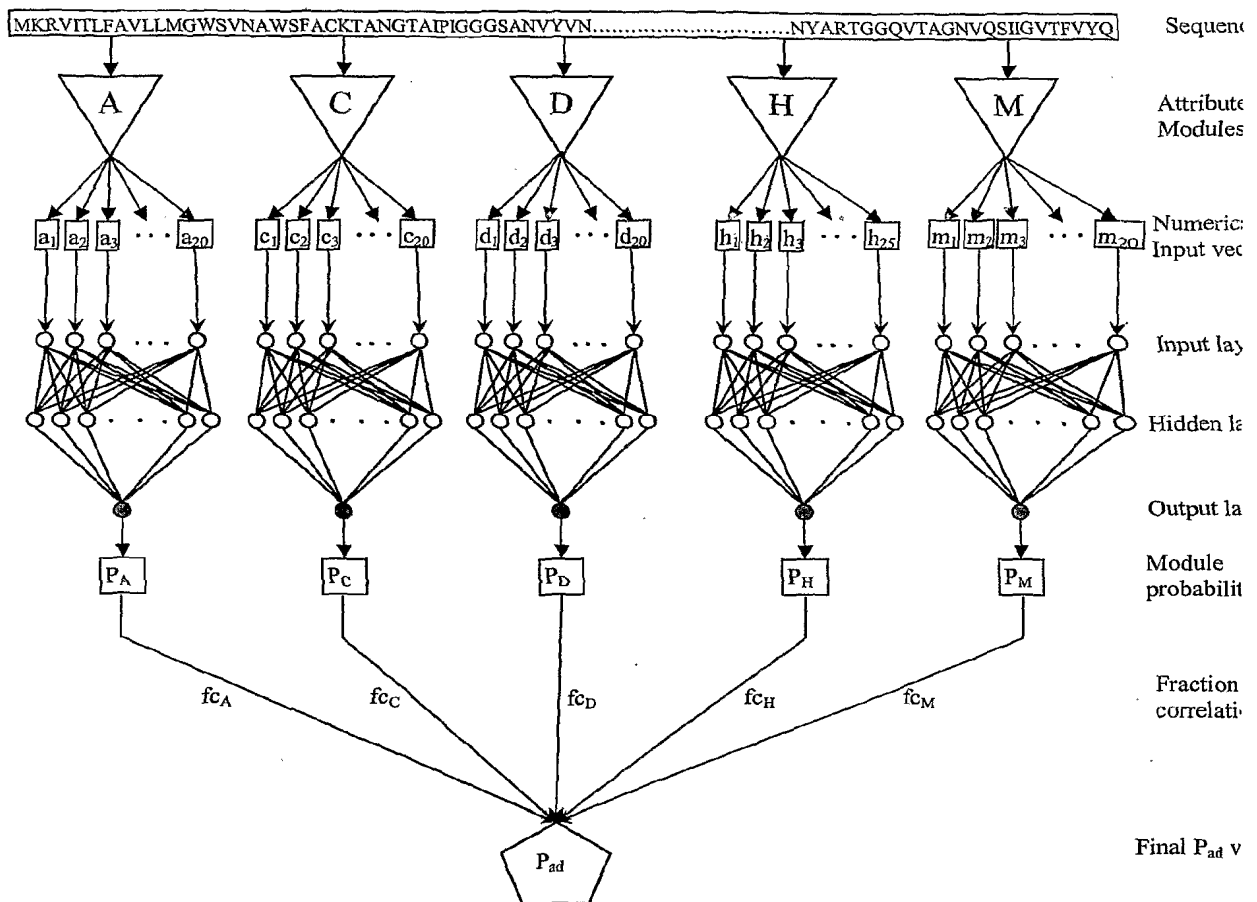
20. A network as claimed in claim 18, wherein the hidden layer comprises of neurons represented as 30 for amino acid frequencies, 28 for multiplet frequencies, 28 for dipeptide frequencies, 30 for charge composition and 30 for hydrophobic composition.

21. A network as claimed in claim 18, wherein the output layer comprises of neurons to deliver the output values as probability value for each protein sequence.

1/5

The Neural Network architecture

Figure 1



Assessment of SPAAN using defined test dataset.

Figure 2

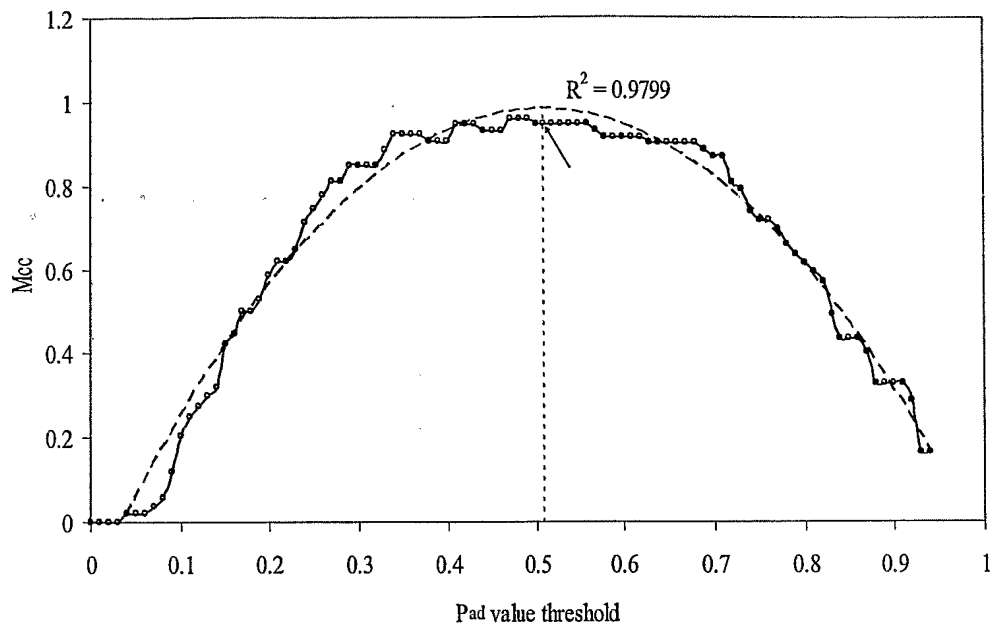


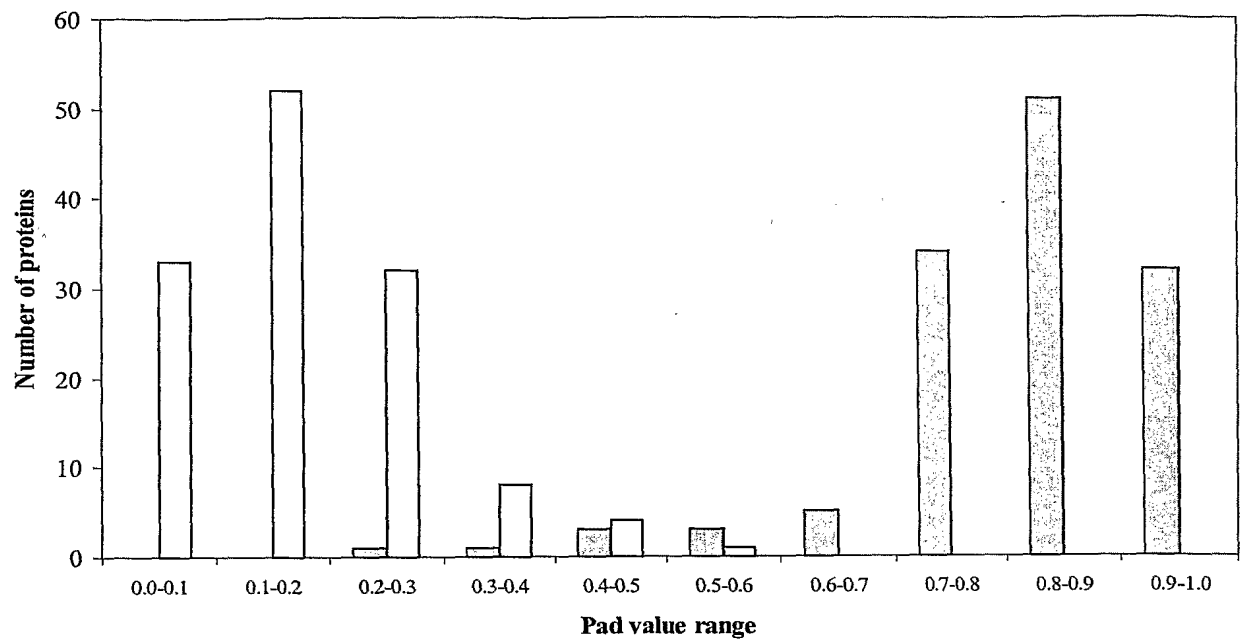
Figure 3 (a)

Figure 3 (b)

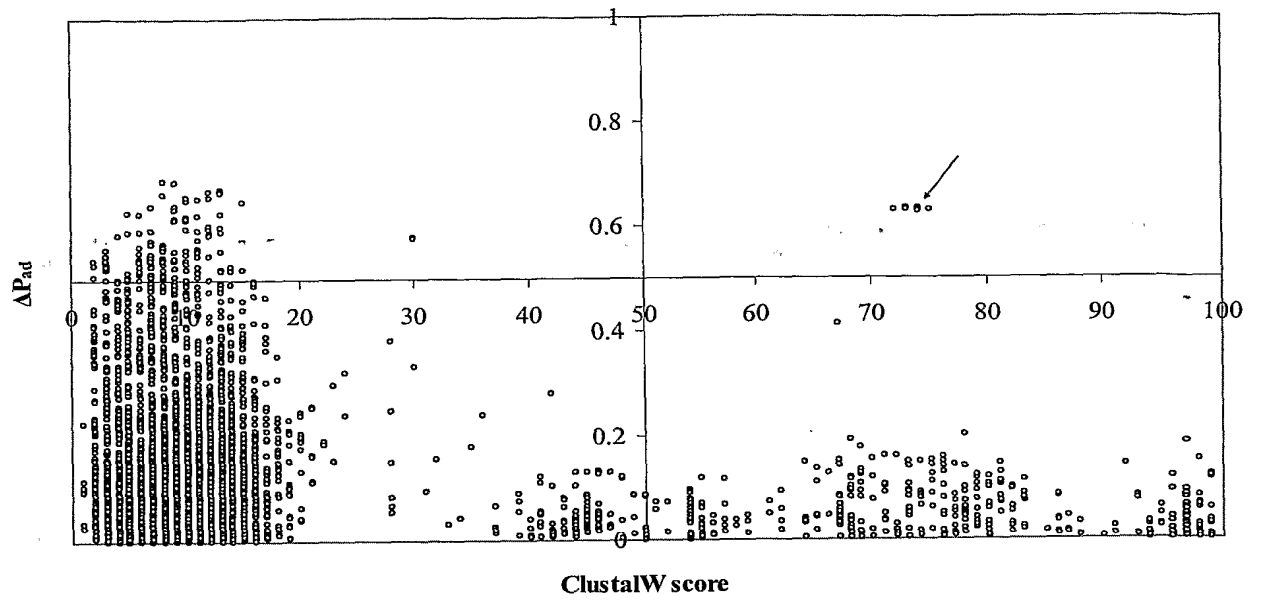
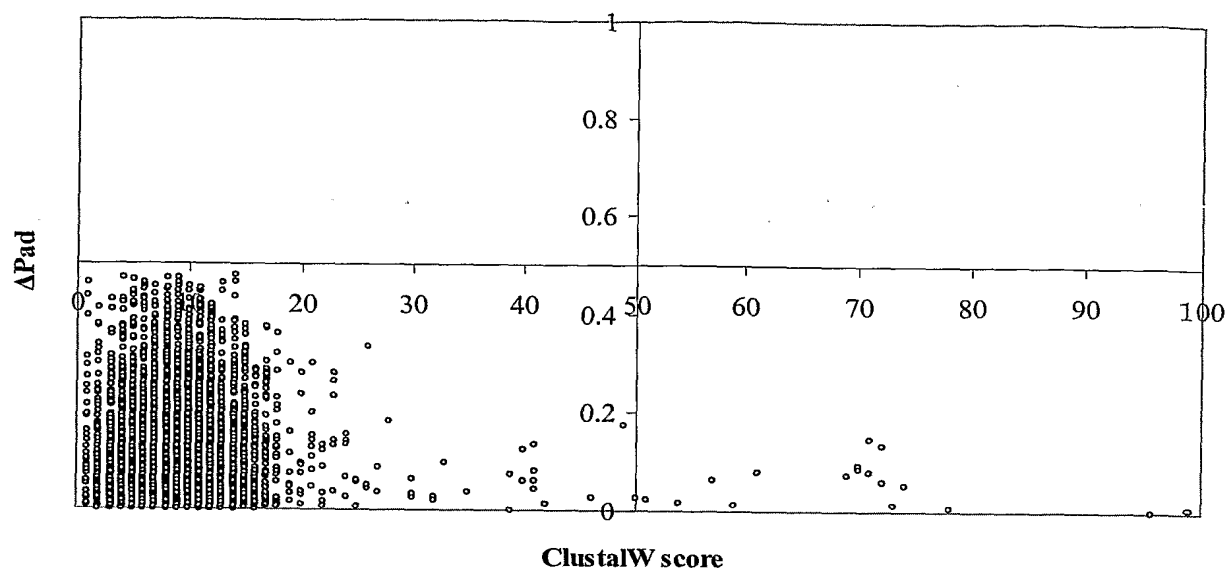


Figure 3(c)



Application Project

<120> Title :

<130> AppFileReference :

5 <140> CurrentAppNumber :

<141> CurrentFilingDate :

Sequence

10 <213> OrganismName : Escherichia coli O157:H7

<400> PreSequenceString :

MINLSKEATV GKALTPIAIL MMLSFPVASQ AAGLVIKNGT VYNANGVPVV DINKPNSSGL 60
 SHNIWDLNLV DKNGVVFNNNS ANESSTSLAG NIQGNNSLTS GSAKVILNEV TSKNPSTING 120
 MMFVLRUKAR LFLANPKCET VNTGGGCHTW KLTLETSPFD IQDDKTAQTS VNGSTITLST 120
 15 LDNASPTEIL SRNVVVNGKV SADELNVVAG NNYVNAAGQV TGSVSATGSR NGYSVDVAKL 240
 GGMVANKISL VSTEKGVGVR NLGVIAGGVN GVSIDSKGNL LNSNAQIQSA STINLTNGT 300
 LDNTTGTVTS VGTISLNTNK NTIVNTRAGN ISTMGDIYVN SGTIDNTNGK LAAAGMLAVD 360
 TNNATLINSK KGSSVGIEAG LVALKTGTLN NSNGQIRGGY VGLESAALNN NNGDIQTGTD 420
 IAIISNGNVD NNGGLIRSST GHIVIGAAGS VNNSTKTAD TGSSDSLGI ADTGVIEGAN 480
 20 NINNNGGQIA SNGNVSLSSY STIDDYAGKI LSNSKVIIKG SSLRNDTGGI SGKQGIEVAV 540
 GGSILTNNIGV ISSEEGDISL LANSVDNHGG FMGQONITME SMSGVNNNTA LIVASKKLKI 600
 NARGDLNKRD DIALSANSVA YFGMPQQTGG MVGKEGIELS GQNIYNNNSR LIAEDGPLTL 660
 QAQNTFDNTR ALVTSGADAS IQVGGTYNN YATTWSAGNL DIDATTLQNS SSGTMIDNNA 720
 TGFIASDKNL SLEVNVSLTN YGWSIGKGDV DVTVNNGNLY NRNTIAAEKG LDIAALNGIE 780
 25 NWKDISAGGD LTMNTNRHVT NNSNSNMVGG NIVINAVNDI NNRGNIVSDA DLNVTTKGNL 840
 YNYLYMVGYG DIALSANSVA NNNATIEATG DLIIDSKGNV GNNRGNLHAL NGVLSVKGN 900
 LNNNGEIRG YGDTVLTALT NYDSYKGS LT SETGDTVLT NIVDNAYGLI AGENVSVDK 960
 STIYNTALI AANKKLIVINA GGNLENRDGN NFLRNNGALF GITDNVGGIV GKEGVTL 1020
 NVYNNSSII AENGPLNLLS RGTLDNTRAL LSSGADAIIR AAGTFYNNYA TTYSAGNL 1080
 30 YAASLNNASD GRLEDNTATG VIASDKNLDL SVDNSVTNYG WISGKGDVHF NVLKGTLYNR 1140
 NAIAADNALT INALNGVENF KDIVAGTALT IDTQKYVTNN SNSNMLGQTI AINAVNDINN 1200
 RGNIVGDYSL GVKTTGNIYN YLNMLSYGVA GVSANKVTNS GKDAVLGGFY GLALEANETD 1260
 NTGTIVGM 1268
 <212> Type : PRT
 35 <211> Length : 1268
 SequenceName : SEQ ID 1
 SequenceDescription :

Sequence

40 <213> OrganismName : Escherichia coli O157:H7

<400> PreSequenceString :

MNTIHLRCLF RMNPLVWCLW ADVAAKLRLS KRYSVFTFQR MKFMNRTSPY YCRRSVLSLL 60
 ISALIYAPPV MAAFTPDVIG VVNDETVDS QRVDERGTN NTHIINHQQQ NVYGGVSN 120
 45 LIESGGYQDV GRHNNYVQS NNTTINGGRQ SIHDGGISTG TIESGNQDV YKGGISNGT 180
 IKGGASRVEG GSANGTLIDG GSQIVKVQGH ADGTTINKSG SQDVVQGS LA TINTINGGRQ 240
 YVEQSTVETT TIKNGGEQVR YESRALDTTI EGGTQSLNSK STAKNTQIYS GGTQIIDNTS 300
 SSDVIEVYSG GVLVDVSGGTA TNVTQHDGAI LKTNNTGTTV SGTNSEGAFS IHNHVADNVL 360
 LENGGHLDIN AYGSANKTII KDKGTMSVLT NAKADATRID NGGVMDVAGN ATNTIINGGT 420
 50 QNINNYGIAT GTNINSGTON IKSGGKADTT IISGSRQVV EKDGTAIGSN ISAGGSLIVY 480
 TGGIAHGVNQ ETGSALVANT GAGTDIEGYN KLSHFTITGG EANYVVENT GELTVVAKTS 540
 AKNTTIDAGG KLIVQKEAKT DSTRLNNGGV LEVDQDGEAK HVEQQSGGAL IASTTSGTLI 600
 EGTNSYGDFA YIRNSEAKNV VLENAGSLTV VTGSRVDTI INANGKMDVY GKDVGTVLNS 660
 AGTQTIYASA TSDKANIKGG KQTVYGLATE ANIESGEQIV DGGSTKTHI NGGTQTVQNY 720
 55 GKAINTDIVS GLQQIMANGT AEGSIINGGS QIVNEGLAE NSVLNDGGTL DVREKGSATG 780
 IQQSSQALV ATTRATRVTG TRADGVAFSI EQGAANNILL ANGGVLTVES DTSSDKTQVN 840
 TGGREIVKTK ATATGTTLTG GEQIVEGVAN ETTINDGGIQ TVSANGEAIK TTINEGGTIL 900
 VNDNGKATDI VQNSGAALQT STANGIEISG THQYGTFSIS GNLATNMLE NGGNLLVL 960
 TEARDSTVGK GGAMQNGQD SATKVNSSGQ YTLGRSKDEF QALARAEDLQ VAGGTAIVYA 1020
 60 GTLADASVSG ATGSLSLMTP RDNVTPVKLE GAIRITDSAT LTIGNGVDTT LADLTAASRG 1080
 SVWLNSNSNC AGTSNCEYRV NSLLLNDGNV YLSAQTAAPA TTNGIYNTLT TNELSGSGNF 1140
 YLHTNVAGSR GDQLVVNNNA TGNFKIFVQD TGVSFQSDDA MTLVKTTGGD ASFSLGNTG 1200
 FVDLGTIEYV LKSDGNSNWN LTNDVKPNPD PNPNNPNPK PDPKPDPKPD PKPDPPTPEPT 1260
 PTPVPEKRIT PSTAAVLNMA ATLPLVFDAE LNSIRERLNI MKASPHNNNV WGATYNTRNN 1320
 65 VTTDAGAGFE QTLTGMTVGI DSPNDIPEGI ATLGAFMGYS HSHIGFDRGG HGSVGSYSLG 1380
 GYASWEHESG FYLDGVVKLN RFESNVAGKM SSGGAANGSY HSNGLGGHIE TGMRTDGN 1440
 NLTPYASLTG FTADNPYHL SNGMESKSVD TRSIYRELGA TLSYNMRLGN GMEIEPWLKA 1500

AVRKEFVDDN RVKVNNDGNF VNDLSGRRGI YQAGIKASFS STLSGHLGVG YSHGAGVESP 1560
WNAVAGVNWS F 1571
<212> Type : PRT
<211> Length : 1571
5 SequenceName : SEQ ID 2
SequenceDescription :

Sequence

10 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MAVKISGVLK DGTGKPVENC TIQLKARNNS ATVVVNTVAS ENPDEAGRYS MDVEYGQYSV 60
ILLVEGFPPS HAGTITVYED SQPGTLNDFL GAMTEDDVVRP EALRRFELMV EEVARNASAV 120
AQNTAAAYKS ACPAGTAAEF AATHATDAAL SAPALMTSAG QAASSAQAS SSAGTASTKA 150
15 TEASKSAAAA ESSKSAAATS AGAAKTSETN AAVSQQSAAT SASTATTKAS EAASSARDAS 240
ASKEAAKSSE TSAASSASSA ASSATAAGNS AKAAKTSETN AKSSETAAEQ SASAAAGSKT 300
AAALSASAAS TSAGQASASA TAAGKSAESA ASSASTATTK AGEATEQASA AASSASAAKT 360
SETNAKASET SAESSKTAAA SSASSAASSA SSASASKDEA TRQASAAKSS ATTASTKATE 420
AAGSATAAAQ SKSTAESAAT RAETAACKRAE DIASVALED ASTTKKGIVQ LSSATNSTSE 480
20 SLAATPKAVK AAYELANGKY TAQDATTAQK GIVQLSNATN STSEMLAATP KSVKAAAYDLA 540
NGKYTAQDAT TAQKGIVQLS SATNSASETL AATPKAVKAA NDNANGRVPS ARKVNKGALS 600
SDITLTPKDI GTLNSTTMSF SGGAGWFKLA TVTMPQASSV VSITLIGGAG FNVGSPQQAG 660
ISELVLRAGN GNPKITGAL WQRTSTGFTN FAWVNTSGDT YDIYVAIGNY ATGVNIQWDY 720
TSNASVTIHT SPAYSANKPE GLTDGTVYSL YTPSEQFYPP GAPIPWPSDT VPSGYALMQG 780
25 QTFDKSAYPK LAAAYPSGVI PDMRGWTIKG KPASGRAVLS QEODGIKSHS HSASASSTD 840
GKTKTSSFDY GTKSTNNTGA HTHSVSGTAA SAGNHHSVT GASAVSQWSQ NGSVHKVVS 900
ASVNTSAAGA HTHSVSGTAA SAGAHHTVG IGAHHSVAI GSHGHTITVN AAGNAENTVK 960
NIAFNYIVRL A 971
<212> Type : PRT
30 <211> Length : 971
SequenceName : SEQ ID 3
SequenceDescription :

Sequence

35 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MKRVITLFAV LLMGWSVNAW SFACKTANGT AIPIGGGSAN VYVNLAPAVN VGQNLVVDLS 60
TQIFCHNDYP ETITDYVTLQ RGSAYGGVLS NFSGTVKYSG SSYPFPTTSE TPRVVYNSRT 120
40 DKPWPVALYL TPVSSAGGVA IKAGSLIAVL ILRQTKNYS DDFQFVWNIY ANNDVVVPTG 180
GCDVSARDVT VTLPDYPGSV PIPLTVYCAK SQNLGYLSG TTADAGNSIF TNTASFSPAQ 240
GVGVQLTRNG TIIPANNTVS LGAVGTSAVS LGLTANYART GGQVTAGNVQ SIIGVTFVYQ 300

<212> Type : PRT
45 <211> Length : 300
SequenceName : SEQ ID 4
SequenceDescription :

Sequence

50 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MGYVTGGLPM KNNRAWALIS GLILFSGTAP AADNLHFTGN LLGKSCTPVI NGNLLAEIHF 60
PTIAASDLMQ RGQSDRVPLV FQLKDCKSTT AFNVKVTLMG TEDTDLPGFL SIDSSSSATG 120
55 VGIGIETAGG AAVPINSTTG ASFPLNQGN SVNFNAWLQT VNGRNVTS GD FTATMTVTFE 180
YF 182
<212> Type : PRT
<211> Length : 182
SequenceName : SEQ ID 5
60 SequenceDescription :

Sequence

65 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MKRHLNTSYR LVWNHITGTL VVASELARSR GKRAVAVAL SLAAVTSVPA LAADKVVQAG 60
ETVNDGTLTN HDNQIVFGTA NGMTISTGLE LGPDSEENTG GQWIQNGGIA GNTTVTTNGR 120

QVVLEGGTAS DTVIRDGGGQ SLNGLAVNTT LNNRGEQWVH EGGVATGTII NRDGYQSVKS 180
GGLATGTIIN TGAEGGPDSD NSYTGQKVQG TAESTTINKN GRQIILFSGL ARDTLIYAGG 240
DQSVHGRALN TTLNGGYQYV HRDGLALNTV INEGGWQVVK AGGAAGNTTI NQNGELRVHA 300
GGEATAVTQN TGGALVTSTA ATVIGTNRLG NFFTENGKAD GVVLES GGRL DVLESHSAQN 360
5 TLVDDGGLA VSAGGKATSV TITSGGALIA DSGATVEGTN ASGKFSIDGT SGQASGLLLE 420
NGGSFTVNAG GQAGNTTVGH RGTLLTAAAG SLSGRTQLSK GASMLVNGDV VSTGDIVNAG 480
EIRFDNQTP NAALSRAVAK SNSPVTFHKL TTTNLTGQGG TINMRVRLDG SNASDQLVIN 540
GGQATGKTWL AFTNVGNSNL GVATTGQGIR VVDAQNGATT EEGAFALSRP LQAGAFNYTL 600
NRDSDEWYL RSENAYRAEV PLYTSMLTQA MDYDRILAGS RSHQTGVNGE NNSVRLSIQG 660
10 GHLGHDNNGG IARGATPESS GSYGFVRLEG DLLRTEVAGM SLTTGVYGAA GHSSVDVKDD 720
DGSRAGTVRD DAGSLGGYLN LVHTSSGLWA DIVAQGTRHS MKASSDNDF RARGWGWLGS 780
LETGLPFSIT DNLMLPEQLQ YTWQGLSLDD GQDNAGYVKF GHGSAQHVRA GFRLGSHNDM 840
TFEGTSSRD TLRDSAKHSV SELPVNWWVQ PSVIRTFSSR GDMSMGTA AAA GSNMTFSPSR 900
NGTSLDLQAC LEARIRENIT LCVQAGYAHS VCGGSAEGYN GQATLNMFT 960
15 <212> Type : PRT
<211> Length : 949
SequenceName : SEQ ID 6
SequenceDescription :
20 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
25 MKKWHYIFCI ILPHLGLPCG YAANDGTCAT RGGHTLSLN FPLTTVSAAN NVPGNLTIDI 60
ANATSSSENY VLNCDSDSKHS NGAYHEIYYT ADPAAGMVYS TTASGLAFYY LNEYVDVGTK 120
ISVLNAGYTA VPFEHVSNOA TTTDHTCQGN KTTAVGVSLK TGADAKISFR IKRSINGTVV 180
IPITDIALLY ANISSTTTTRG EAIKVRISG SLTAPQSCQI NAGQVIYFDF DTIPASEFSS 240
TAGQAITSRK ITKTVSIECT GMYERTQKV DASFTGTNRS SDDTMVATDN ADVGIKIYNK 300
SNAEVSNNNG KLPADMGNTT IFGRKNGSVT FSAAPASFTG ARPQPGVFNA TATLTIEFVN 360
30 <212> Type : PRT
<211> Length : 360
SequenceName : SEQ ID 7
SequenceDescription :
35 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
40 MSRYKTGHKQ PRFRYSVLAR CVAWANISVQ VLFPLAVTFT PVMAARAQHA VQPRLSMGNT 60
TVTADNNVEK NVASFAANAG TFLSSQPDSD ATRNFITGMA TAKANQEIQE WLGKYGTARV 120
KLNVDKDFSL KDSSLEMLYP IYDPTNMLF TQGAHRTDD RTQSNIGFGW RHFSGNDWMA 180
GVNTFIDHDL SRSHTRIGVG AEYWRDYLKL SANGYIRASG WKKSPDIEDY QERPANGWDI 240
RAEGYLPAPW QLGASLMEYQ YYGDEVGLFG KDKRQKDPHA ISAEVTTYTPV PLLTSLAGHK 300
45 QKSGGENDTR FGLEVNRYRIG EPLAKQLDTE SIRERRVLAG SRYDLVERN NIVLEYRKSE 360
VIRIALPERI EGKGGQTLSL GLVVSKATHG LKNVQWEAPS LLAEGGKITG QGSQWQVTL 420
AYRPGKDNYY AISA VAYDNK GNTSKRVQTE VVITGAGMSA DRTALTLDGQ SRIQMLANGN 480
EQKPLVLSLR DAEGQPVGTG KDQIKTELTF KPAGNIVTRS LKATKSQAKP TLGEFTETEA 540
GVYQSFTTG TQSGEATITV SVDGMSKTIV AELRATMMDV ANSTLSANEP SGDVVADGQ 600
50 AYTLTLTAVD SEGPNVTGEA SRLRFVPQDT NGVTVGAISE IKPGVYSAV SSTRAGNVVV 660
RAFSEQYQLG TLQOTLK FVA GPLDAAHSSI TLNPDKPVVG GTVTAIWTVK DAYDNPVTS 720
TPEAPSLAGA AAEGSTASGW TNNGDGTWTA QITLGSTAGE LEVMPKLNQ NAAANA AKVT 780
VVADALSSNQ SKVSV AEDHV KAGESTTVTL VAKDAHGNAI SGLALSASLT GTASEGATVS 840
SWTEKNGSY VATLTTGGKT GELRVMLPFN GQPAATEAAQ LTVIAGEMSS ANSTLVADNK 900
55 APTVKTTTEL TFTVKDAYGN PVTGLKPDAP VFSGAAGTGS ERPSAGNWTE KNGGVVSTL 960
TLGSAAGQLS VMPRVNGQNA VQPLVLNVA GDASKAEIRD MTVKVNNQLA NGQSANQITL 1020
TVVDTYGNPL QGQEVTLTL VQVTSKTGNT VTTNAAGKAD IELMSTVAGE HNISASVNGA 1080
QKTVTVKFNA DASTGQANLQ VDAQAQVAN GKDAFTLTAN VEDKNGNPVP GSVLTFNLPR 1140
GVKPLTGDNV WVKANDEGKA ELQVVSVTAG TYEITASAGN SQPSNTQTIT FVADKATATV 1200
60 SGIEVIGNYA LADGNAKQTY KVTVT DANNN LLKDSEVTLT ASPANLVLT NGTAKTNEQG 1260
QAIFTATTTV AAKYTLTAKV SQADQGSEIS TAESKFVADD TNAVLTASSD VTSLVADGIS 1320
TAKLEVTLMS ANNPVGGNMW VDIKTPEGVT EKDYQFLPSK NDHFVSGKIT RTFSTSKPGV 1380
YTFTFNALTY GGYEMKPVTV TITAVDADTA KGEEAMN 1417
65 <212> Type : PRT
<211> Length : 1417
SequenceName : SEQ ID 8
SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
5 MARGWASSEA SGAMTDWLNN FGTARISLGV DEDFSLKNSQ FDFLHPWYDT PDYLLESQHT 60
LHRTDDRTQI NTGLGWRHFT SSWMSGINLF FDHDLTRYHS RAGLGAEYWR DYLLKSSNAY 120
IGLTGWRSAE ELDNDFEARP ANGWDLRAEG WLPAPWQLGG KLVYEQYYGD EVALFDKNDR 180
QSNPHAITAG LNYTPFPPLLT LSAEQRQKQK GENDTRFAVD LTWQPSSSMQ KQLNPDEVAG 240
10 RRSLAGSRYD LIDRNNNIVL EYRKKELRRL SLLDPVKGKS GEIKPLVSSL QTKYALKGYN 300
IEAAALEAAG GKVSTSGKDI TVTLPGYRFT NTPETDNTWS IDVTAEDVKG NLSRHEQSMV 360
VIQAPTLQSK DSSLNVNPLT VAADKKSTTT LTVTAHSDSG TPVPGLALQT RSEGVQDITL 420
SDWTNNGDGS YTOILTATTT SGSVTLTPQI NGESAVKESI VVNIVPVVSS RDHSSITIDN 480
VSYVAGNDIK VVVELKDSSN QFVAYQKEEL VKAVTVENSK PGATIVWHEE QPGVYAANYF 540
15 AYKQGTALRA QLSLHNWNP LQSHIYNIEA NQNKARVATL SATNNDVYAD KKTFTNLTIN 600
VTDESDNPLT NHQVTFKNEK GSAEFVEPPQ QNTDAYGVAT INMVSQVAEE NTISATLPNG 660
FSQRIIAKFV SDSSTPKFKQ LVADPDITIA GNSQGSTLTA IITDFHNNPL KDMKVNFPVAP 720
GGSQDLNNTA TTDQSGIVRV HLTSSKAGSY SVDASLEVDK NIHQSVTITV VPNREQSVMT 780
LNAGSGSAIA NNTNIVTLTA SVKDVYGHPL PDEDVKFTLP ASMTGNFTLS SETARTDANG 840
20 DAVVTLRGTK AGEFTVTATL TRNNTVAYQQ VTFIGDTNSA QLQPLTASLN SIVAGNSTGS 900
TLTATILDAY QNPLKQDLVT PQSNDVTLSE TEVTNTLTGQ ATVTMTSNIA GQHNVVVSRK 960
AQASDNGTFP VSVLDPDESSA KVISITGAEK TITRDTQSGQV TATITSTLAE TLTNVVQVPG 1020
AQPTTNITIG DTAYTDNNGY AYVNLSTQP GVVQVTATLD NNSSSKVDVN VANGKLELTS 1080
SKPETTVHNS EGITLTATAR NARGELMPGQ IITFSVTPEG ATLSNTGEVL TDQSGQAKVT 1140
25 LTSDKVNVTY VTAIMGKDVV VQSQVTVAVK ADAKTAHVVS VVASPDITTA DGIDSSTITS 1200
RVEDDYGFVP RVEDDISHGLD TKGSPVNVIP TITRDTQSGQV TATITSTLAE TLTNVVQVPG 1260
TANQSATITL VAGTADESKE ILKSDVDTLK ADYQQSAKLT LTLQDKYGNP IVTSDHLEFV 1320
QSGPFVNFLE LSDIDYSQRN YGEYTVTVTG GKEGTATLIP MLNGVHQANL SISLNLIQSI 1380
KEMSGHVIAN NHTFSTAKFP SEGFAAYYT LNNDNFEAGK TVDDYMFSSS QGWVSVDSAG 1440
30 KVSFANIGDQ VSVTISAVPR QGGTTYQTLI KLKGGWVWNG NHTNIWLAAN ALCHAKNDGY 1500
NLPGITHTLS GENKRTQGS L YGEWGNVGAF SSNSQFTPGA YWTSESDDYS RHYYVQMLTG 1560
MTGSDADSSP QLTACRKS L 1579
<212> Type : PRT
<211> Length : 1579
35 SequenceName : SEQ ID 9
SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
40 MITHGCYTRT RHKHKLLKKT L IMLSAGLGLF FYVNQNSFAN GENYFKLGSD SKLLTHDSYQ 60
NRLFYTLKTG ETVADLSKSQ DINLSTIWSL NKHLYSSESE MMKAAPGQQI ILPLKKLPFE 120
YSALPLLGS A PLVAAGGVAG HTNKLTKMSP DVTKSNTDD KALNYAAQQA ASLGSQQLSR 180
45 SLNGDYAKDT ALGIAGNQAS SQLQAWLQHY GTAENVLQSG NNFDGSSLDL LPPFYDSEKM 240
LAFGQVGARY IDSRFTANLG AGQRFFLPAN MLGYNVFIQ DFSGDNTRLG IGGEYWRDYF 300
KSSVNGYFRM SGWHESYNKK DYDERPANGF DIRFNGYLP YPALGAKLIY EQYYGDNVAL 360
FNSDKLQSNP GAATVG VNYT PIPLVTMGID YRHGTGNEND LLYSMQFRYQ FDKSWSQQIE 420
PQYVNELR TL SGRYDLVQR NMNIILEYK QDILSLNIPH DINGTEHSTQ KIQLIVKSKY 480
50 GLDRIVWDDS ALRSQGGQIQ HSGSQSAQDY QAILPAYVQG GSNIYKVTAR AYDRNGNSSN 540
NVQLTITVLS NGQVVDQGVG TDFTADKTS A KADNADTITY TATVKKNGVA QANVPVSFNI 600
VSGTATLGAN SAKTDANGKA TVTLKSSTPG QVVVSATKAE MTSALNASAV IFFDQTKASI 660
TEIKADKTTA VANGKDAIKY TVKVMKNGQP VNNQSVTFST NFGMFNGKSQ TQATTGNDGR 720
ATITLTSSSA GKATVSATVS DGAEVKATEV TFFDELKIDN KVDIIGNNVR GELPNIWLYQ 780
55 GQFKLKASGG DGTYSWYSEN TSIA TVDASG KVTLNGKGSV VIKATSGDKQ TVSYTIKAPS 840
YMIKVDKQAY YADAMSICKN LLPSTQTVLS DIYDSWGAAN KYSHYSSMNS ITAWIKQTSS 900
EQRSGVSSTY NLITQNPLPG VNVNTPNVYA VCVE 934
<212> Type : PRT
<211> Length : 934
60 SequenceName : SEQ ID 10
SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
65 MLVLSESFKN KLLPMNGYMK GGSDSGSKAQ ARATEKGIEL QREMWQTNMQ NLAPFTPLAQ 60

QYVSQQLNLS SLQGGQALN QYYSQQYKD LAGQARYQSL AAAEATGGLG STATGNQLAA 120
IAPTLGQNWL SGQMNNYNNL ANIGLGALTG QANAGQNYAN NVSQLYQQQA AASAAANKP 180
SGLQSFATGA IGGAASGAMI GSAVPVIGTG IGALAGGVIG GLGSLF 226
<212> Type : PRT
5 <211> Length : 226
SequenceName : SEQ ID 11
SequenceDescription :

Sequence

10 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MKKILSGLIL LLCCPYGFAA NGDGATHMSN LSFGLPTVAA ANNHSGYNIF EALSNTTGTY 60
PVRCHCDDTH SGPGQQTAF FIFYTGDAAAP GLVLERTLNG LNYALNDYL SVGVTFIFIIN 120
15 NQYAAIPFEH LSNQSTSPQH TCGAGNNGST VNLDSGRSAK LSFYVRHSIT GTVTIPTTEV 180
AWLYAGMSDH FPKTTPVSKV TIRGQLTAPQ NCELTPNQSI DVDFQKINSA EFSSTAGSII 240
AERKIKTEVT VSCGTMEDVR STEVVSASMI AANRSADATM IVTSNPDVGI KIPDKNDRPV 300
NVDGGLPAD MGAISRLGKT DGSVTFYSAP ASLTGAKPAP DNGFTATATL VIEFTN 356

20 <212> Type : PRT
<211> Length : 356
SequenceName : SEQ ID 12
SequenceDescription :

Sequence

25 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MNKIYRLKWN RSRNCWSVCS ELGSRVKGKK SRAVLISAIS LYSSLVFADD VIVNQDKTID 60
30 FGKENQSDY RITVTDNANL VINATDTSRP RLTLASGGGL DITGGKVTIN GPLNFLKGT 120
GFLNVSNAGS ELYADDLYES NSGMRHGRGY FNVSNGGKIHK VKGTSRLTYL QGNVSGEGSQ 180
VNSETFFMGV YGSYGGNQYL SVNNGGEVNA RKQISLGYYD QVSDTTLAVS EGGKISAPTI 240
SLSTNSLAL GAQEGSAKA AGIIDAIEKIE FVWAKTSEKK ITLNHTDKDA TISADIVSGS 300
EGLGYINALN GTTYLTGDNS AFSGKVKIEQ NGALGITQNI GTAEINNRGK LHLKADDSMT 360
35 FANKISGNCT ISIDSGTVEL TGNNAFSGY IDVASGAVAV ISEDKNIGRA ELDVDGKLQI 420
NANKDWVFDN DLEGRGIVEI NMGNHEFSFD EFAYTDWFQG SLAFQNTTFN LEKNAEFLQK 480
GGITAGQSSL VTVGKAHSI STLGFSGGTV DFGALTAGAQ MTEGTVNVSK TLDLRGEGVI 540
QVSDSDVVRV VSRDIDSALS LTEVDDGNSI IKLVDAQGAE VLGDAGNLQL QDKNGQILSS 600
SAQRDIQNG QKAAVGTYDY RLTSQVNNND LYIGYGLTQL DLHATDSDAL VLSSNGKSEN 660
40 AADLSAKITG SGDLAFSSQK GQTVSLSNKD NDYTGVTDLR SGTLLLNNDN VLGNTHLRL 720
AAETELDMNG HSQTGVTLNG SADSLLSLNG GSLTVTNGGT STGSLTGSSE LNIQGGTLDI 780
AGDNSNLNAN VNIAANSANVL VSHAQGLGSA NVENNGTLAL NNSAEKRAAA SVNYALGGNL 840
TNNGTLMTGM SGQQAGNVLV VKGNYHGNG QLVMTVNLNG DDSVTDKLVV EGDTSCTTAV 900
TVNNAGTGGA KTLNGIELIH VDGKSEGEFV QAGRIVAGAY DYTTLARGQGA NSGNWYLTSG 960
45 SDSPELQPEP DPMNPPEPNP NPEPNPNPTP TPGPDLNVDN DLRPEAGSYI ANLAAANTMF 1020
TTRLHERLGN TTYTDMVTGE QKQTTMWMRH EGGHNKWRDG SGQLKTQSNR YVLQLGGDVA 1080
QWSQNGSDRW HVGVMAGYGN SDSKTISSRT GYRAKASVNG YSTGLYATWY ADDESNGAY 1140
LDSWAQYSWF DNTVKGDDLQ SESYKSKGFT ASLEAGYKHK LAEFNGSQGT RNEWYVQPQA 1200
QVTWMGVKAD KHRESNGTLV HSNCGDGNVQT RLGVKTWLKS HHKMDDGKSR EFQPFVEVNW 1260
50 LHNSKDFSTS MDGVSVDQDG ARNIAEIKTG VEGQLNANLN VWGNVGVQVA DRGYNDTSAM 1320
VGIKWQF 1327
<212> Type : PRT
<211> Length : 1327
SequenceName : SEQ ID 13
SequenceDescription :

Sequence

60 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MITMKKSVLT AFITVVCATS SVMAADDNAI TDGSVTFNGK VIAPACTLVA ATKDSVVTLP 60
DVSATKLQTN GQVSGVQTDV PIELKDCDIT VTKNATFTFN GTADTTQITA FANQASSDAA 120
TNVALQMYMN DGTIAIKPDT ETGNILLQDG DQTLTFKVDY IATGKATSGN VNAVTFNHIN 180
YY 182
65 <212> Type : PRT
<211> Length : 182
SequenceName : SEQ ID 14

SequenceDescription :

Sequence

5 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MSKFVKTAIA ATMVMGAFAS TSTIAAGNNG TARFYGTIED SPCSIVPDDH KLEVDMGDIG 60
SGILKNNGTS TPKAFQIHLQ DCVFDQTMM TTTFTGNASS TNSGNYTYIY NTDTGAAFNN 120
VSLAIGDAQG TSYKSGAGIE QKIVNDTATN KGKAKQTLDF KAWLVGAADA PDLGNFEANT 180
10 TFQITYL 187
<212> Type : PRT
<211> Length : 187
SequenceName : SEQ ID 15
SequenceDescription :

15

Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
20 MRVIFLRKEY LSLLPSPMIAS LFSANGVAAA IDLCQGYDIK ASCHASRQSL SGITQVWSIA 60
DGQWLVSFDM TNNASGGAVF LQQGAFTLS PENETGMTLF ANNTVSGEYN NGGAI FAKEN 120
STLNLTDVIF SGNVAGGYGG AIYSSGTNDT GAIDLRTVNA VFRNNIANDG KGGAIY TINN 180
DIYLSDDVFN NNQAYTSTSY SDGDGGAIDV TDNNSDSKHP SGYTIINNTA FTNNTAEGYG 240
GAIYTN SATA PYLIDISVDD SYSQNGGVLV DENNSAAGYG DGPSSAAGGF MYLGLSEVTF 300
25 DIADGKTLVI GNTENDGA VD SIAGTGLITK TSGSDLVLNA DNNDFTGEMQ IENG EVTLGR 360
SNSLMNVGDT HCQDDPQDCY GLTIGSIDKY QNQAELNVGS TQQTFAHSLT GFQNGTLNID 420
AGGNVTVNQG SFAGTIEGAG QLTIAQNGSY VLAGAQSMAL TGDIVVDAGA VLSLEGDAAD 480
LAALQDDPQS IVLNGGMLDL SDFSTWQSGT SYKDGLEVSQ SSGTVIGSQD VVDLAGGNM 540
HIGGDGKDG V YVVIDAGDGQ VSLANDNQYL GTTQIASGTL MVSDNSQLGY THYNRQVIFT 600
30 DKPQESVMEI TANVDTRSTT TEHGRDIEMR ADGEVAVDAG VDTQWGALMA DSSGQHDEG 660
STLTKTGAGT LELTAGSTTQ SAVRVEEGTL QGVADIFPY ASSLWVGDA TFFTADQDI 720
QSIDATSSGT IDISDGTVLR LTGQDTSVAL NASLFNC DGT LVNATDGVTL TGELNTNLET 780
DSLTYLSNVT VNGNLNTSG AVSLQNGVAG DTLTVNGDYT GGGLLLDSE LNGDSDVSDQ 840
LVMNGNTAGN TTVVNSITG IGEPTSTGIK VVDFAADPTQ FQNNAQFSLA GSGYVNMGAY 900
35 DYTLVEDNND WYLRSEQVTP PSPDPDPTP DPDPDPTP TPDPEPTPAY QPVLNAKVGG 960
YLNLRANQ AFMMERRDHA GGDGQTLNLR VIGGDYHYTA AGQLAQHEDT STVQLSGDLF 1020
SGRWGTDGEW MLGIVGGYSD NQGDSRSSMT GTRADNQNHG YAVGLTSSWF QHGKQKQAW 1080
LDNWLQYAWF SNDVSEHEDG VDHYHSSGII ASLEAGYQWL PGRGVVIEPQ AQVIYQGVQQ 1140
DDFTAANRAR VSQSQDDIQ TRLGLHSEWR TAVHVIPTLD LNYHDPHST EIEEDASTIS 1200
40 DDAVKQRGEI KVGVTGNISQ RVSLRGVAV QKGSDDFAQT AGFLSMTVKW 1250
<212> Type : PRT
<211> Length : 1250
SequenceName : SEQ ID 16
SequenceDescription :

45

Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
50 MHSWKKKLIV SQLALACTLA ITSQANAATN DISGQTYNTF HHYN DATYAD DVYYDGYVGW 60
NNYAADSYN GDIYPVINNA TVNGVISTYY LDDGISTNTN ANSLTIKNT IHGMITSECM 120
TTDCADDRAT GYVYDRLTSL VDNSTIDDNY EHYTYNGTYN NAADTHVVDV YDMGTAITLD 180
QEVDSLITNN SHVAGITLTQ GYEWEDIDDN TVSTGVNSSE VFNNTITVKD STVTSGSWTD 240
EGTTGWFGHT GNASYSNTL TADDVAIAAI ANPYADNAMQ TTVTLDNSTL MGDVVFSSNF 300
55 DENFFPQGAN SYRDADGDVD TNGWDGTD RM DVTLNNGSKW VGAAMSVHMV DEDGDGSYDG 360
YAVGTEATAT LLDIAANSLW PSSTVGVDNI NTQYDENGHI VGNEVYQSGL FNVTLNGGSE 420
WDTTKSSLID TLSINSGSQV NVADSRLLSD TVSLTGGSNL NIGEDGHVAT NTLTIDNSTV 480
KMSDDVSAGW GLEDAALYAN TITVTNDGLL DINVDQFDAN PFQADTLNLT STTD TNNGNIH 540
AGVFDIHSSD YVMDTDLVND RTNDTTKSNY GYGLIAMNSD GHLTINGNGD NDNTASIEAG 600
60 QNEVDNNGDH VAAATGNYKV RIDNATGAGS IADYNGNELI YVNDKNSNAT FSAANKADLG 660
AYTYQAEQRG TTVVLQQMEL TDYANMALSI PSANTNIWNL EQDVTGTRLT NSRHGLADNG 720
GAWVS YFGGN FNGDNGTINY DQDVNGIMVG VDKIDGNNA KWIVGAAAGF AKGDMNDRSG 780
QVDQDSQTAY IYSSAHFANN VFVDGSLSYS HPMNDLSATM SNGTYVDGST NSDAWGFLK 840
AGYDFKLGA GYVTPYGSIS GLFQSGDDYQ LSNMVKVDGQ SYDSMRVELG VDAGYFTTYS 900
65 EDQALTPYFK LAYVYDSSNN DNDVNGDSID NGTEGSAVRV GLGTQFSFTK NFSAYTDANY 960
LGGGDVDQDW SANVGKYTW 980
<212> Type : PRT

<211> Length : 980
SequenceName : SEQ ID 17
SequenceDescription :

5 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MKLKHVGMIV VSVLAMSSAA VSAAEGDESV TTTVNGGVIH FKGEVVNAAC AIDSESMNQT 60
10 VELGQVRSSR LAKAGDLSSA VGFNIKLNDC DTNVSSNAAV AFLGTTVTSN DDTLALQSSA 120
AGSAQNVGIQ ILDRITGEVLI LDGATFSAKT DLIDGTNILP FQARYIALGQ SVAGTANADA 180
TFKVQYL 187
<212> Type : PRT
<211> Length : 187
15 SequenceName : SEQ ID 18
SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MKLLKVAATA AIVFSGSALA GVVVPQYGGGG GNHGGGGNNS GPNSSELNIYQ YGGGNSALAL 60
20 QADARNSDLT ITQHGGGNGA DVGQGSDDSS IDLTQRGFGN SATLDQWNGK DSHMTVKQFG 120
GGNGAAVDQT ASNSTVNVTQ VGFGNNATAH QY 152
<212> Type : PRT
<211> Length : 152
25 SequenceName : SEQ ID 19
SequenceDescription :

30 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MPIGNLGHNP NVNNSIPPAP PLPSQTDGAG GRGQLINSTG PLGSRALFTP VRNSMADSGD 60
35 NRASDVPLP VNPMLAASE ITLNDGFEVL HDHGPLDTLN RQIGSSVFRV ETQEDGKHIA 120
VGQRNGVETS VVLSDQYAR LQSIDPEGKD KFLVTGGRGG AGHAMVTVAS DITEARQRIL 180
ELLEPKGTGE SKGAGESKGV GELRESNSGA ENTETQTST STSSLRSDPK LWLALGTVAT 240
GLIGLAATGI VQALALTPEP DSPTTDPDA AASATETATR DQLTKEAFQN PDNQKVNIDE 300
LGNAPSGVL KDDVVANIEE QAKAAGEEAK QQAIENNAQA QKKYDEQQAK RQEELKVSSG 360
40 AGYGLSGALI LGGGIGVAVT AALHRKNQPV EQTTTTTTTT TTTSARTVEN KPANNTPAQQ 420
NVDTPGSEDIT MESRRSSMAS TSSTFFDTSS IGTVQNPYAD VKTSLHDSQV PTSNSNTSVQ 480
NMGNTDSVVY STIQHPPRDT TDNGARLLGN PSAGIQSTYA RLALSGGLRH DMGGLTGGSN 540
SAVNTSNNPP APGSHRFV 558
<212> Type : PRT
45 <211> Length : 558
SequenceName : SEQ ID 20
SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MFSTFKKAAL LAAIALPFST MAAPTIVTFQG EVTDTQTCVN INGQNSVVL MPTVAMADFG 60
ATLADGQSAG QTPFTVSVSN CQAPTQADQA INTTFLGYDV DASTGVMGMR DTSSDAAKGF 120
55 GIQLMDSSTS GNPVTLGAT NVPGLTLKVG DTEASYDFGA RYFVIDSAAA TAGKITAVAE 180
YTLSYL 186
<212> Type : PRT
<211> Length : 186
60 SequenceName : SEQ ID 21
SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
65 <400> PreSequenceString :
MNSEGGKPGN VLTVNGNYTG NNGLMTFNAT LGGDNSTPK MNVKGDTQGN TRVRVDNIGG 60
VGAQTVNGIE LIEVGNSAG NFALTGTGVE AGAYVYTLLAK GKGNDKKNWY LTSKWDGVTP 120

ADTPDPINNP PVVDPEGPSV YRPEAGSYIS NIAAANSLSF HRLHDLGEP QYTDSLHSQD 180
SASSMWMRHV GGHERRSSAGD GQLNTQANRY VLQLGGDLAQ⁺ WSSNAQDRWH LGVMAGYANQ 240
HSNTQSNRVG YKSDGRISGY SAGLYATWYQ NDANKTGAYV DSWALYNWFD NSVSSDNRSA 300
DDYDSRGVTA SVEGGYTFFEA GTCSGSEGTL NTWYVQPQAQ ITWMGVKDSH HARKDGTREI 360
5 TEGDGNVQTR LGVKTYLNSH HQRDDGKQRE FQPYIEANWI NNSKVYAVKM NGQTVSRDGA 420
RNLGEVRTGV EAKVNNNLSL WGNVGVQLGD KGYSDTQGM L GVKYSW 466
<212> Type : PRT
<211> Length : 466
SequenceName : SEQ ID 22
10 SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
15 <400> PreSequenceString :
MSYLNLRLYQ RNTQCLHIRK HRLAGFFVRL FVACAFVQA PLSSAELYFN PRFLADDPQA 60
VADLSRFENG QELPPGTYRV DIYLNNGYMA TRDVTFTNGD SEQGIVPCLT RAQLASMGLN 120
TASVAGMNL ADDACVPLTT MVQDATAHLD VGQQRNLNLT PQAQFMSNRAR GYIPPELWDP 180
GINAGLLNLYN FSGNSVQNRI GGNSHYAYLN LQSGNLIGAW RLRDNTTWSY NSSDRSSGSK 240
20 NKWQHINTWL ERDIIPLRSR LTLGDGYTQG DIFDGINFRG AQLASDDNML PDSQRGFAPV 300
IHGIARGTAG VTIKQNGYDI YNSTVPPGPF TINDIYAAGN SGDLQVTIKE ADGSTQIFTV 360
PYSSVPLLQR EGHTRYSLTA GEYRSGNAQQ EKPRFFQSTL LHGLPAGWTI YGGTQLADRY 420
RAFNFVGIGKN MGALGALSVD MTQANSTLPD DSQHDGQSVR FLYNKSLNES GTNIQLVGYR 480
YSTSGYFNFA DTTYSRMNGY NIETQDGVIO VKPKFTDYNN LAYNKRGLQ LTVTQQLGRS 540
25 STLYLSGSHQ TYWGTSNVDE QFQAGLNTAF EDINWTLNYS LTKNAWQKGR DQMLARNVNI 600
PFSHWLRSDS YSMHDLNLR MTNLAGVYGT LLEDNNLSYS VQTGYAGGGD 660
GNSGSTGYAT LNYRGGYGNA NIGYSHSDDI KQLYYGVSGG VLAHANGVTL GQPLNDTVVL 720
VKAPGAKDAK VENQTVGRTD WRGYAVLPYA TEYRENRVAL DTNTLADNVD LDNAVANVVP 780
TRGAIVRAEF KARVGKILM TLTHNNKPLP FGMVTSSESS QSSGIVADNG QVYLSGMPLA 840
30 GKVQVKWGE ENAHCVANVQ LPPESQQQLL TQLSAECR 878
<212> Type : PRT
<211> Length : 878
SequenceName : SEQ ID 23
SequenceDescription :

35 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
40 MQIIFGEKCV SLLRLFFAAV LMLWCAQTAA YSQQCHTTQG NPYIGVNFV KTLLEEEENTT 60
GVVKDKFYQW NESNDYYVSC DCDKDNVRSR RWAFAADSPL VYLDGNWYKI NDYLAQKVL 120
QVKGSSPTAV PFENVGTGAD TRWHICDPGG QRLGGQGASG NSGSFSLKIL QPFVGSVIP 180
PMALARLFEC YNIPAGDSCT TTGTPVLVYY LSGTINSLGS CSVNAGETIE VDLGDVFAAN 240
FRVVGHKPLG ARTAELAIPV RCNTGNAGLV NVNLSLTATT DPSYPQAIKT SRPGVGVVVT 300
45 DSQNNIISPA GGTLPPLSIPD DADSIA 326
<212> Type : PRT
<211> Length : 326
SequenceName : SEQ ID 24
SequenceDescription :

50 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
55 MKIKTLAIVV LSALSLSSTA ALAAATTVNG GTVHFKEGVV NAACAVDAGS VDQTVQLGQV 60
RTASLAQDGA TSSAVGFNIQ LNDCTNVAS KAAVAF LGTV IDAGHTNVLA LQSSAAGSAT 120
NVGQILDRY GAALTLDGAT FSEQTTLNNG TNTIPFQARY YAIQEATPGA ANADATFKVQ 180
YQ 182
<212> Type : PRT
60 <211> Length : 182
SequenceName : SEQ ID 25
SequenceDescription :

65 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :

MKLKVIATLI ATVAVGVSFN SNFASASTTS ASLTVNSNLT MGTCSAQIMD NSNKVINEVV 60
 FGNVYISELG AKSKVQQFKI RFSNCSGLPQ NSAQIVLAPN GISCAGSQSS SAGFSNKFTD 120
 ASAATRTAVE VWTDTTPESN GSTQFHCAQK IPVPVTLPAD TTTQPYDYPL SARMTVAEGR 180
 LVTDVDPGNF RSPPTFTITY Q 201

5 <212> Type : PRT
 <211> Length : 201
 SequenceName : SEQ ID 26
 SequenceDescription :

10 Sequence

 <213> OrganismName : Escherichia coli O157:H7
 <400> PreSequenceString :

15 MASTVEYGET VDGVLLEKDI QLVYGTANNT KINPGSEQHI KEFGISSNTE INGGYQYIEM 60
 NGTAEYSVLN DGYQIVQMGG AANQTTLNNG VLQVYGAAND PTIKGGRLLV EKDGITVLAA 120
 IEKGGLLEVK EGGLAIAVDQ KAGGAIKAST RVMEVFGTNR LGQFEIKNGI ANNMLLENGG 180
 SLRVEENDFA YNTTVDSGGL LEVMDGGTAT GVDKKAGGKL IVSTNALEVS GTNSKGQFSI 240
 KDGVSKNYEL DDGSGLIVME DTQAIDTILD EHATMQSLGK DTGTRVQANA VYDLGRSDQN 300
 GSITYSSKAI SENMVINNGR ANVWAGTMVN VSVRGNDGIL EVMKQPINYA PAMLVGKVVV 360
 20 SEGASLRTHG AVDTSKADVS LENSATWIIA DITTTNQNTN LNLANLAMSG ANVIMMDES 420
 TRSSVTASAE NFFTLLTNTL SGNGNFYMRM DMANHQSDDL NVTGQATGDF KIFVTDGAS 480
 PAAGDSLTLV TTGGGDAAPT LGNAGGVVDI GTYBYTLLDN GNHSWSLAEN RAQITPSTTD 540
 VLNMAAAQPL VFDAELDTVR ERLGSVKGVS YDTAMWSSAI NTRNNVTDA GAGFEQTLTG 600
 LTLGIDSRFS REESSTIRGL FFGYSHSDIG FDRGGKGNVD SYTLGAYAGW EHONGAYVDG 660
 25 VVKVDRFANT IHGKMSNGAT AFGDYNNGA GAHVESGFRW VDGLWSVRPY LAFTGFTTDG 720
 QDYTLNGMR ADVGNTRLR AEAGTAVSYH MDLQNGTTLE PWLKAAVRQE YADSNQVKVN 780
 DDGKFNNDDVA GTRGVYQAGI RSSFTPTLSG HLSVSYNGA GVESPWNTQA GVVWTF 836

<212> Type : PRT
 30 <211> Length : 836
 SequenceName : SEQ ID 27
 SequenceDescription :

35 Sequence

 <213> OrganismName : Escherichia coli O157:H7
 <400> PreSequenceString :

40 MQRKGNKLLI QLCSVILLFF TTSWYALANE CYIERNAEGD YHMKISSTQL SLASQMVEVP 60
 TEIAEATWDV NIQLRGDAIG CKSLGDSKAV HFLNTADPSL ISTYTTTNGA ALLKTTVPPI 120
 VYSVELLCLS CGAADELDLW LPAQSGADNF IPSTQTKWAY EYSDQSWYLR FRLFITPEFK 180
 PKNGVSSGTT IAGKIASWYI GTNDQPWINF YIDNDSLKFF VDEPTCATVA LAQDQGNVSG 240
 NQVTLGNSYV SEVKNGLTRE IPFSIRAAYC YASKITVKLK AANKPSDATL VGKTTGSASG 300
 VAVKVNSTYD NSKVLLKADG SNTVDYNFAA WSNLLFLPF TAQLVPDGSQ NAVGVGTFSG 360
 NATFSFTYE 369

45 <212> Type : PRT
 <211> Length : 369
 SequenceName : SEQ ID 28
 SequenceDescription :

50 Sequence

 <213> OrganismName : Escherichia coli O157:H7
 <400> PreSequenceString :

55 MYQFTHQKSR IPKKTLLAAC CALFYSSNGA AADTVEYDSS FLMGTGASTI DVKRYAQGNP 60
 TPPGLYNVRV FVNGQATSSL EIPFVDIGEN SAAACLTHKN LAQLHIKQPE QPVTLLAREG 120
 EEEDCLDLAK SYEKADVCFD GSDQFLDLTI PQAYVLKSYG GYVDPSSLWES GINAATLAYT 180
 LNAHYTSSDN DNSDSVYGAF NSGINLGAWH FRARGYNWT TDNGSDFDQ DRYLQRDIPA 240
 IRSQIIMGDA YTTGETFDSV NVRGVRLYSD SRMLPSALAS YAPTIRGVAN SNAKVTVQS 300
 GYKIYETTVP PGEFVIDDIS PSGFGSELVV TIEADGSKR TFTQPFSSVV QMQRPGVGRW 360
 60 DFSAGKVIDD SLRSEPMMGQ ASYYYGLNNL FTGYTGIIQT DNNYLALGLL VGINTSIGAF 420
 AVDVTHSRAE IPDDKTYQGQ SYRVTNWKLK QDTGTSFNLA AYRYSTQDYL GLHDALVLID 480
 DAKHLSADED KNTMQTYSRM KNQFTVSINQ PLNIAYEDYG SLFISGSWTY YWAANNRTE 540
 YNVGYSKSVS WGSFVSNLQR SWNEDGEKDD AMYVSVSVPI ENILGGKRKS SGFRNLNTQL 600
 NTDFDGSQHL NVNSSGNTEN NLVNVSVNAG YSLDKNAGDL ASVGGYLYNE SGLGGISASA 660
 65 SATSDNSQQY SISTDGGFVL HSGGLTFTNN SFSSNDTLVL INALGAKGAR INNSNNEIDR 720
 WGYAVTSSVS PYRENVRGLN IETLENDVEL KSTSATTVPR SGSVVLTRFE TDEGRSAVLN 780
 ITAANGKSIP FAAEVYQGEV MIGSMGQGGQ AFVRGINDSG ELIVRWYENN QTIDCKLHYQ 840

FPAQPQTQGS TNTLLLNLT CQVANH 866
<212> Type : PRT
<211> Length : 866
SequenceName : SEQ ID 29
SequenceDescription :

5 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
10 MKFKRLLHSG IASLSLVACG VNAATDLGPA GDIHFSITIT TKACEMEKSD LEVDMGMTML 60
QKPAAVGTVL SKKDFTI ELK ECDGISKATV EMDSQSDSDD DSMFALEAGG ATGVALKIED 120
DKGTQQVPKG SSGTPIEWAI DGETTSLHYQ ASYVVVNTQA TGGTANALVN FSITYE 176

15 <212> Type : PRT
<211> Length : 176
SequenceName : SEQ ID 30
SequenceDescription :

20 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
25 MKYNNIIFLG LCLGLTTYSA LSADSVIKIS GRVLDYGCTV SSDSLNFTVD LQKNSARQFP 60
TTGSTSPAVP FQITLSECSK GTTGVRVAFN GIEDAENNTL LKLDEGSNTA SGLGIEILDG 120
NMRPVKLN DL HAGMWI PLV PEQNNILPYS ARLKSTQKSV NPGLVRASAT FTLEFQ 176

30 <212> Type : PRT
<211> Length : 176
SequenceName : SEQ ID 31
SequenceDescription :

Sequence

35 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MKWRKRGYLL AAILALASAT IQAADVTITV NGKVVAKPCT VSTTNATVDL GDLYSFSLMS 60
AGAASAWHDV ALELTNCPVG TSRVTASFSG AADSTGYKN QGTAQNIQLE LQDDSGNTLN 120
TGATKTVQVD DSSQSAHFPL QVRALTVNGG ATQGTIQAVI SITYTYS 167

40 <212> Type : PRT
<211> Length : 167
SequenceName : SEQ ID 32
SequenceDescription :

45 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
50 MKRAPLITGL LLISTSCAYA SSEGCCADST SGATNYSSVV DDVTVNQTDN VTGREFTSAT 60
LSSTNQYAC SCSAGKAVKL VYMVSPVLTT TGHQTGYKYL NDSLDIKTMN RPNPND 117

55 <212> Type : PRT
<211> Length : 117
SequenceName : SEQ ID 33
SequenceDescription :

Sequence

60 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MKKALLAAL VMASGSALAV DGGHIDFNGM VQSGTCKVGV VDTGMHSTTT DGVVTLDTAN 60
VTDTFAEVSA TAVGLLPKEF MISVECDPGA PKNAELTMGS ASYANTSGTL NNNMNITVNG 120
IAPQNVNIA VHNMKNKAGA AEIKQVHMNN SSEVQELTLD AEGKGQYVFN ASYVKAPNSP 180
AVTAGHVTN ALYTVAYK 198

65 <212> Type : PRT
<211> Length : 198
SequenceName : SEQ ID 34

SequenceDescription :

Sequence

5 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MKPNMIVGAL ALTSVFMAGH LQAADGTVHF RGEIIDSTCE VTPETKDQVV DLGKVNRTAF 60
SGVDDVAAPT AFSIDLTQCP ETFKSAAIRF DGNEDAHNGG NLAIGTPLDN SNDAAAGISP 120
SDNSGDYTGGA GAVSAAKGVA IRLYNRADNT QVKLYENSAS TPISNGNASM KFMARYIATE 180
10 TTIDPGTANA DSQFTVEYIK 200
<212> Type : PRT
<211> Length : 200
SequenceName : SEQ ID 35
SequenceDescription :

Sequence

20 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MPIFQREGHL KYSFAAGEYQ AGNYDSASPR FGQLDLIYGL PWGMTAYGGV LISNNYNAFT 60
LGIGKNFGYI GAISIDVTQA KSELNNDRDS QGQSYRFLYS KSFESGTDPR LAGYRYSTSG 120
FYTFQBATDV RSDADSDYNR YHKRSEIQGN LTQQLGAYGS VYLNLTQQDY WNDAGKQNTV 180
SAGYNGRIGK VSYSIAYSWN KSPWDESDR LWSFNISVPL GRAWSNYRVT TDQDGRTNQQ 240
VGVSGTLLED RNLSYSVQEG YASNGVGNNG NANVGQGGG GNVNVGYSYG KDYRQLNYSV 300
25 RGGVIVHSEG VTLSQLGET MTLISVPGAR NARVVNNGGV QVDWMGNAIV PYAMPYRENE 360
ISLRSDSLGD DVDVENAFQK VVPTRGAIVR ARFDTRVGYR VLMTLLRSAG SPVPFGATAT 420
LITDKQNEVS SIVGEEGQLY ISGMPEEGRV LIKWGNDASQ QCVAPYKLSL ELKQGGIIPV 480
SANCQ 485
<212> Type : PRT
30 <211> Length : 485
SequenceName : SEQ ID 36
SequenceDescription :

Sequence

35 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MSGYTVKPPT GDSNEQTQFI DYFNLFYSKR DQEQISISQQ LGNYGATFFS ASRQSYWNTS 60
RSDQQISFGL NVPFGDITTS LNYSYSNNIW QNDRDHLLAF TLNVPPFSHW RTDSQSAFRN 120
40 SNASYSMSND LKGGMTNLSG VYGTLLPDNN LNYSVQVGNT HGGNTSSGTS GYSTLNRYGA 180
YGNTNVGYSR SGDSSQIYYG MSGGIIAHAD GITFGQPLGD TMVLVKAPGA DNVKIENQITG 240
IHTDWRGYAI LPFATEYREN RVALNANSLA DNVELDETVV TVIPTHGAIA RATFNAQIGG 300
KVLMTLKYGN KSVPPGAIVT HGENKNGSIV AENGQVYLTG LPQSGKLQVS WGNDRKNSNCI 360
VDYKLPEVSP GTLLNQQTAI CR 382
45 <212> Type : PRT
<211> Length : 382
SequenceName : SEQ ID 37
SequenceDescription :

Sequence

50 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MSALYERSQL TQVMISSAPA TAETMEKAIEY LRLDCTIKEV QFTAGQKQDI DVTTLCTSTEQ 60
55 ENINGLGASS EISMSGNFYL NQAQNALRDA YDNDTVYAFK VQFPGSGKGFK FLAEVRQHTW 120
SSGTNGVVAA TFSRLRLKGP VSYVPLAFV KNLDKTLTVN TGALLTMSVS VNGGTPPYKH 180
AWKKDGPVE GQTTDTFSKP GAQSGDKGAY TCEVTDSAEQ PQSITSDACT VTVNGAGG 238
<212> Type : PRT
60 <211> Length : 238
SequenceName : SEQ ID 38
SequenceDescription :

Sequence

65 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :

MRNKPFFYLLC AFLWLAVSHA LAADSTITIR GYVRDNGCSV AAESTNFTVD LMENAAKQFN 60
NIGATTPVVP FRILLSPCGN AVSAVKVGFT GVADSHNANL LALENTVSAA SGLGIQLLNE 120
QQNQIPLNAP SSAISWTTLT PGKPNTLNFY ARLMATQVPV TAGHINATAT FTLEYQ 176

5 <212> Type : PRT
<211> Length : 176
SequenceName : SEQ ID 39
SequenceDescription :

10 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
15 MNKSVVSISA AMLVLLCQPV MGSEISPATP SDEDNYTTFDP QLFRGSRFSQ SSLAKLITRE 60
SVAPGNKYMD IYTNKLSGS WNVTFKEAAD GRVLPCLTPE VADAIGLKTG EDKGEKDPVC 120
TFAKELAPGI TSQTQLSQLR LDLSVPQSQL ISRPRGYVPP SELDTGASLA FMNYIANYYN 180
VAYSGQNAHS QRSWASFNG GINLGAWQYR QLSNMTWDND KGNQWNNIRS YLQRPLPAIN 240
SQLMMGLIT SGRFFSGLSY HGVSLATDER MLPDSMRGYA PTIRGVAATN ARVSVMQNGH 300
EIIYQTTVAPG PFEINDLYPT SYSGDLDTVT TEANGAVSRF SVPFSAVPES MRPGTSRYNV 360
20 EVGKTQDSGD DSMFGDLTWQ HGMTNTLTFTN SGSRIADGYQ ALMLGGVYGS SLGAFGANLT 420
WSHARVPES AQSGWMSQLT WSKTFQPTST TVSLAGYRYS TSGYRDLADV LGERHAASNK 480
QSWDSSQWRQ QSRFDLTLSQ SLANYGNLFV SGSTQNYRGG KSRDTQLQLG YSNSFSHGIS 540
MNL SVGRQRM GGYKDNSDDM QTVTSLSFSF PLGGNGPRVP SLSNSWTHST DGSSQLQSSL 600
TGMLDEAQT NYSLNVMRDQ QYKQTTLSGN MQKRFSQTTV GLNASKGQDY WQASGNVQGA 660
25 MAVHGGGITF GPYLGETFAL VEAKGAEGAK VYNSSQLEIN DSGYALVPAV TPYRYNRISL 720
DPQGMDSGAE LVDSEKQVAP VAGAAVKVIF RTRPGKALLI KSRMADGSEL PMGADVLDEN 780
NTVVGIAQGQ GQIYLRTETQ KGHLSVRWGE GANDSCQLPF DISGKDSNSP IIRLNETCQS 840

30 <212> Type : PRT
<211> Length : 840
SequenceName : SEQ ID 40
SequenceDescription :

35 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
40 MKLAACFLTL LPGAFAVAASW TSPGFPAFSE QGTGTFVSHA QLPKGTRPLT LNFDQQCWQP 60
ADAIKLNQML SLQPCSNTPP QWRLFRDGKY TLQIDTRSGT PTLMISIQNA AEPVANLVRE 120
CPKWDGLPLT LDVSATFPEG AAVRDYYSQQ IAIKNGQIT LQPAATSNGL LLLERAETDA 180
SAPFDWENAT VYFVLTRDFE NGDPSNDQSY GRHKDGMAEI GTFHGGDLRG LTNKLDYLQQ 240
LGVNALWISA PFEQIHGWG GGTGKDFPHY AYHGYTQDW TNLDANMGNE ADLRTLVDNA 300
HQRGIRILFD VVMNHTGYAT LADMQEYQFG ALYLSGDEVK KTLGERWSDW KPAAGQWHS 360
FNDYINFSDK TGWDKWWGKN WIRTDIGDYD NPGFDDLTMS LAFLPDIKTE STTASGLPVF 420
45 YKNKTDTHAK AIDGFTPRDY LTHWLSQWVR DYGIDGFRVD TAKHVELPAW QQLKTEASAA 480
LREWKANPD KALDDKPFWM TGEAWGHGVM QSDYYRHGFD AMINFQYQEQ AAKAVDCIAQ 540
MDTTWQQMAE KLQGFNVLSY LSSHDTLRLR EGGDKAAELL LLAPGAVQIF YGDESSRPFG 600
PTGSDPLQGT RSPMNWQDVS GKSAANVAHW QKISQFRARH PAIGAGKQTT LSLKQGYGFV 660
REHGDDKVLV IWAGQQ 676

50 <212> Type : PRT
<211> Length : 676
SequenceName : SEQ ID 41
SequenceDescription :

55 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
60 MPQRHHQGHK RTPKQLALII KRCLPMVLTG SGMCLTTANA EBYFDPIML ETTKSGMQTT 60
DLSRFSKYA QLPGTYQVDI WLNKKKVSQK KITFTANAEQ LLQPQFTVEQ LRELGIKVDE 120
IPALAEKDD SVINSLEQII PGTAAEFDNF HQRNLNLSIP IALYRDARGY VSPSRWDDGI 180
PTLFTNYSFT GSDNRYRQGN RSQRQYLNMQ NGANFGPWRL RNYSTWTRND QASSWNTISS 240
YLQRDIKALK SQQLLGESAT SGSIFSSYNF TGVQLASDDN MLPNSQRGFA PTVRGIANSS 300
AIVTIRONGY VIYQSNVPAG AFEINDLYPS SNSGDLEVTI EESDGTQRRF IQPYSSLPM 360
65 QRPGLHKYSA TAGRYRADAN SDSKEPEFAE ATAIYGLNNT FTLYGGLLGS EDYYALGIGI 420
GGTLGALGAL SMIDNRADTQ FDNQHSFHY PETNTNIAVS YYRYTNDGYF 480
SFDEANTRNW DYNRQKSEI QFNISQTIFD GVSLYASGSQ QDYWGNNEKN RNISVGVSGQ 540

QWIGIGYSLNY QYSRYTDQNN DRALSLNLST PLERWLPRSR VSYQMTSQKD RPTQHEMRID 600
GSLDDGRLS YSLEQSLDDD NNHNSSVNAS YRSPYGTFS A GYSYGNDSSQ YNYGVTTGGVV 660
IHPHGVTLST YLGNALALID ANGASGVRIQ NYPGIATDPF GYAVVPYLT YQENRLSVDT 720
TQLPDNDVLE QTTQFVVPNR GAMVAARFNA NIGYRVLVTV SDRNGKPLPF GALASNDDTG 780
5 QQSIVDEGGI LYLSSGSSKS QSWTVRWGNQ ADQQCQFAFS TPDSEPTTSV LQGT AQCH 838

<212> Type : PRT
<211> Length : 838
SequenceName : SEQ ID 42
10 SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
15 <400> PreSequenceString :
MMFRNRILLI FILWANFTWA GCRTTASLNI TDGINVGEIL ANETSFSKSV VFTGISCDTS 60
TDKIVYKNIQ SDWVEVGPFG NGEKLVKIE SLGKTS DTIG KSSNAQAVLP YVVKIARGTP 120
DPTGERKSTW FISDTVIANI GGESSSIDF WLGICKALKF NWCVNLYTSK LAGDTFTLGL 180
NISYYPKNTT CKPENTVIKV DDIALFQLRN QGKIAANSKE GTITLKCDNL FGDKKQASRN 240
20 MVVYLSSSDL VKGSNTILRG KTDNGVGFVL DLTEPPKGTE AAIKISANGD QGAATSLWKT 300
DKPGVSLNSN IINIPVMASY VYDEKKVKS GALEATALIN VKYD 344

<212> Type : PRT
<211> Length : 344
SequenceName : SEQ ID 43
25 SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
30 <400> PreSequenceString :
MIKKASLLTA CSVTAFAAWA QDTSPDTLVV TANRFEQPRS TVLAPTTVVT RQDIDRWQST 60
SVNDVLRRLP GVDITQNGGS QQLSSIFIRG TNASHVLVLI DGVRNLNLAGG SGSADLSQFP 120
IALVQRVEYI RGRPSAVYGS DAIGGVNII TTRDEPGTEI SAGWGSNSYQ NYDVSTQQQL 180
GDKTRVTLLG DYATHGYDV VAYGNTGTQA QPDNDGFLSK TLYGALEHNF TDAWSGFVRG 240
35 YGYDNRTNYD AYYSPGSPLV DTRKLYSQSW DAGLRYNGEL IKSQLITSYS HSKDYNYPH 300
YGRYDSSATL DEMKQYTVQW ANNIIIGHGN VGAGVDWQKQ STAPGTAYVK DGYDQRNTGI 360
YLTGLQQVGD FTFEGAARSD DNSQFGRHGT WQTSAGWEFI EGYRFIASYG TSYKAPNLGQ 420
LYGFYGNPNL DPEKSKQWEG AFEGLTAGVN WRISGYRNDV SLDIDYDDHT LKYYNEGKAR 480
IKGVEATANF DTGPLTHTVS YDYVDARNAI TDTPLLRRAK QQVKYQLDWQ LYDFDWGIT Y 540
40 QYLGTRYDKD YSSYPYQTVK MGGVSLWDLA VAYPVTSHLT VRGKIANLFD KDYETVYGYQ 600
TAGREYTLTG SYTF 614

<212> Type : PRT
<211> Length : 614
SequenceName : SEQ ID 44
45 SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
50 <400> PreSequenceString :
MKNKLLFMML TILGAPGIAA AAGYDLANSE YNFAVNELSK SSFNQAIIIG QAGTNNSAQL 60
RQGGSKLLAV VAQEGSSNRA KIDQTGDYNL AYIDQAGSAN DASISQGAY NTAMIIQKGS 120
GNKANITQYG TQKTAIVVQR QSQMAIRVTQ R 151

<212> Type : PRT
55 <211> Length : 151
SequenceName : SEQ ID 45
SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
60 <400> PreSequenceString :
MNIFAYLLVL VFSMSMSSSA FASVMTGTR IIFPGDAKEK TIQLRNTSDQ PYIINIHVED 60
ERGS DKNVFP MPTPTFRME AAAGQALRL L YTGNL PQDR ES VFWFSFSQ LPYLNKNDKS 120
65 QNQLILALT N RVKIFYRPSS IVGKSSDAPK NLTYQVKQNR IEV TNPTGY V TIRAAELLN 180
NGKKVPLANS VMIAPQSTTE WTLPSGISVA PGAQIHLVTV NDYGVNVTSE HAL 233

```
<212> Type : PRT
<211> Length : 233
      SequenceName : SEQ ID 46
      SequenceDescription :
5
Sequence
-----
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
10 MKRLHKRFLI ATFCALLTAT LQAADVITIV NGRVVAKPCT IQTKEANVNI GDLYTRNLQQ      60
   PGSASGWHNI TSLTDCPAE TSAVTAIVTG STDNTGYKYK EGTAENIQIE LRDDQDQATLK      120
   NGDSKTVIVD EITRNAQFPL KARAITVNGN ASQGTIEALI NVIYTWQ      167
<212> Type : PRT
<211> Length : 167
15      SequenceName : SEQ ID 47
      SequenceDescription :

Sequence
-----
20 <213> OrganismName : Escherichia coli O157:H7
   <400> PreSequenceString :
   MRAKLLGIVL TTPAISSFA STETLSFTPD NINADISLGT LSGKTKERVY LAEEGGRKVS      60
   QLDWKFNNAA IIKGAINWDL MPQISIGAAG WTLGSRGGN MVDQDQDSS NPGTWDESR      120
   HPDTQLNYAN EFDLNIKGWL LNEPNYRLGL MAGYQESRYS FTARGGSYIY SSEEGRDDI      180
25   GSFPNGERAI GYKQRFKMPY IGLTGSYRYE DFELGGTFKY SGWVEASDND EHYDPGKRIT      240
   YRSKVKDQNY YSVSVNAGYY VTPNAKVYVE GTWNRVTNKK GNTSLYDHND NTSDYSKNGA      300
   GIENYNFITT AGLKYTF      317
<212> Type : PRT
<211> Length : 317
30      SequenceName : SEQ ID 48
      SequenceDescription :

Sequence
-----
35 <213> OrganismName : Escherichia coli O157:H7
   <400> PreSequenceString :
   MFFKRGKILS AGRLNKKSILG IVMLLSVGLL LAGCSGSKSS DTGTYSGSVY TVKRGDTLYR      60
   ISRTTGTSTVK ELARLNGISP PYTIEVGQKL KLGGAKSSSS TRKSTAKSTT KTASVTPSSA      120
   VPKSSWPPVG QRCNLWPTTG KVIMPYSTAD GGNKGIDISA PRGTPPIYAAG AGKVYVVGNO      180
40   LRGYGNLIMI KHSEDIYITAY AHNDTMLVNN GQSVKAGQKI ATMGSTDAAS VRLHFQIRYR      240
   ATAIDPLRYL PPQGSKPKC      259
<212> Type : PRT
<211> Length : 259
45      SequenceName : SEQ ID 49
      SequenceDescription :

Sequence
-----
50 <213> OrganismName : Escherichia coli O157:H7
   <400> PreSequenceString :
   MPTPNPLAPV KGAGTTLWVY NGNGDPYANP LSDNDWSRLA KVKDLTPGEL TAESYDDSYL      60
   DDEDADWAAT GQGQKSAGDT SFTLAWMPGE QGQALLAWF NEGDTLAYKI RFPNGTVDVF      120
   RGWVSSIGKA VTAKEVITRT VKVTNVGRPS MAEDRSTVTA ATGMTVTPAS TSVVKGQSTT      180
   LTVAFAQPEGA TDKSFRAVSA DTKKATVSVS GMTITVKGVA AGKVNIPVVS GNGEFAAVAE      240
55   INVTAS      246
<212> Type : PRT
<211> Length : 246
      SequenceName : SEQ ID 50
      SequenceDescription :

60 Sequence
-----
   <213> OrganismName : Escherichia coli O157:H7
   <400> PreSequenceString :
65   MSALYERSQL TQVMISSAPA TAETMDKAEY LRLDCTIKEV QFTAGQKQDI DVTTLCSREQ      60
   ENINGLGASS EISMSGNFYL NQAQNALRDA YDNDALYAFK VLFPSGKGFK FLAEVRQHTW      120
   SSGTNGVVAA TFSRLKGGP VSFVVPLAFV KNLDKTLTVN TGALLTMSVS ANGGTPPYKY      180
```

AWKKDGPVD GQTTDTFSKP GAQSADAGKY TCVVTDSAEK AQSVTSVECT VTVSAAAG 238

<212> Type : PRT
<211> Length : 238
5 SequenceName : SEQ ID 51
SequenceDescription :

Sequence

10 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MKKSTLALVV MGIVASASVQ AAEIYNKDGN KLDVYGKVKV MHYMSDND SK DGDQSYIRFG 60
FKGETQINDQ LTGYGRWEAE FAGNKAESDT AQQKTRLAFA GLKYKDLGSF DYGRNLGALY 120
DVEAWTDMFP EFGGDSSAQT DNFMTKRASG LATYRNTDFF GVIDCLNLTL QYQGYNENRD 180
15 VKKQNGDGFG TSLTYDFGGS DFAISGAYTN SDRTNEQNLO SRGTGKRAEA WATGLKYDAN 240
NIYLATFYSE TRKMTPTTGG FANKTQNF EA VAQYQDFDGL RPSLGYVLSK GKDIEGIGDE 300
DLVNYIDVGA TYYFNKNMSA FVDYKINQLD SDNKLNNINND DIVAVGMTYQ F 351

<212> Type : PRT
20 <211> Length : 351
SequenceName : SEQ ID 52
SequenceDescription :

Sequence

25 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MRVKHAVVLL MLISPLSWAG TMTFQFRNPN FGGPNPNGAF LLNSAQANNS YKDPSYNDFF 60
GIETPSALDN FTQAIQSQIL GGLLSNINTG KPGRMVTNDY IVDIANRDGQ LQLNVTDRKT 120
30 GQTSTIQVSG LQNNSTDF 138

<212> Type : PRT
<211> Length : 138
SequenceName : SEQ ID 53
SequenceDescription :

35 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
40 MKRKVLAMLV PALLVAGAAAN AAEIYNKDGN KLDLYGKVVAG LHYFSDDASS DGDMSYARIG 60
FKGETQIADQ FTGYGQWEFN IGANGPESDK GNTATRLAFA GFGFGQNGTF DYGRNYGVVY 120
DVEAWTDMPL EFGGDTYAGA DNFMNNGRANS VATYRNNNGFF GQVDGLNFAL QYQGNNEKSG 180
LFDQEGSGNG NGRKLAKENG DGSVCPLPMT LTLV 214

<212> Type : PRT
45 <211> Length : 214
SequenceName : SEQ ID 54
SequenceDescription :

Sequence

50 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MNTVTLEGGT FNNNGTLNDV VKIEKNSNAV INNTGSLSTL QLHDGTVNNS GIASARVNAQ 60
GDAVFNNLAG GEARKGAILY NSAVVNNAGT WKMGYQDENN NAGTLDIDDK STFNNSGKLI 120
55 LDNSKNAIRF QGSNANATLY NTGEMTLDAAL LGAGAILYDD GASEFINKGV VDAKVTVAVS 180
TAGATESDAF LWNQDGGVIN FDKDNASAVK FTHNNYVALN DGVMNISGNN AVAMEGDKNA 240
QLVNNGVINL GTEGTTDTGL TGMQLDANAT ADAVIENNGT INIFANDSFA FSVLGTEGHI 300
VNNGTIVVIAD GVTGSGLIKQ GDSVNVEGVN GNSGNNTVEVH YTDYTLPDMP NTYTTSPFSE 360
TTDSGSSDGS SNNLNGYIVG TNVDGSAGKL KVNNASMNGV GINTGFAAGT ADTTSVFDNV 420
60 VEGINLTAD AITSTSVVWT AKGSTDASGN VDVIMSKNAY TDVATDASVN DVAKALDAGY 480
TNNELYTSLN VGTAEALNSA LKQVSGSQAT TVFREARVLS NRFSMLADAA PKVGNGLAFN 540
VVAKGDPRAE LGNNTEYDML ALRKTVDLSE SQSMSLEYGI ARLDGDGAQK AGDNGVTGGY 600
SQFFGLKHQM SFDNGMRWNN ALRYDVHNL DSSRSVAYGDV SKTADTDVKQ QYLELRSEGA 660
KTFEPREGLK ITPYAGVKLR HSLEGGYQER NAGDFNLMSN SGSETAVDSI VGLKLDYAGK 720
65 GGSANATLE GGPNLSSYSKS QRTASLAGAG SQHFNVDGQ KGGGINSLAS VGVKYSKES 780
SLNLDAYHWK EDGISDKGVM LNFKKTF 807

<212> Type : PRT

<211> Length : 807
SequenceName : SEQ ID 55
SequenceDescription :

5 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MLNGISNAAS TLGRQLVGIA SRVSSAGGTG FSVAPQAVRL TPVKVHSPFS PGSSNVNART 60
10 IFNVSSQVTS FTPSRPAPPP PTSGQASGAS RPLPPPIAAL KEHLAAYEKS KGPEALGFKP 120
ARQAPPPPTS GQASGASRPL PPIAQALKEH LAAYEKS KGP EALGFKPARQ APPPPTSGQA 180
SGASRPLPPI AQALKEHLAA YEKS KGP EALGFKPARQ APPPPTSGQA 240
AYEQSKKG 248
<212> Type : PRT
15 <211> Length : 248
SequenceName : SEQ ID 56
SequenceDescription :

Sequence

20 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MNKKIHSLAL LVNLGIYGVA QAQEPDTPV SHDDTIVVTA AEQNLQAPGV STITADEIRK 60
NPVARDVSEI IRTMPGVNLT GNSTSGQRGN NRQIDIRGMG PENTLILIDG KPVSSRNSVR 120
25 QGWRGERDTR GDTSWVPPEM IERIEVLRGP AAARYGNAA GGVVNIITKK GSGEWHGSWD 180
AYFNAPEHKE EGATKRTNFS LTGPLGDEFS FRLYGNL DKT QADAWDINQG HQSARAGTYA 240
TTLPAGREGV INK DINGVVR WDFAPLQSL E LEAGYSRQGN LYAGDTQNTN SDA YTRSKYG 300
DETNRLYRQN YSLTWNGGWD NGVTTSNWVQ YEHTRSRIP E GLAGGTEGK FNEKATQDFV 360
DNDLDDVMLH SEVNLPIDFL VNQTLTLGTE WNQQRMKDLS SNTQALTGTN TGG AIDGVSA 420
30 TDRSPYSKAE IFSLFAENNM ELTDSTIVTP GLRFDHHSIV GNNWSPALNI SQGLGDDFTL 480
KMGIARAYKA PSLYQTNPNY ILYSKGQGCY ASAGGCYLQG NDDLKAETSI NKEIGLEFKR 540
DGWLAGITWF RNDYRNKIEA GYVAVGQNAV GTDLYQWDNV PKAVVEGLEG SLNVPVSETV 600
MTWNNITYML KSENKTTGDR LSIIPEYTLN STL SWQARED LSMQTFTWY GKQ QPKKYN Y 660
KGQPAVG PET KEISPYSIVG LSATWDVTKN VSLTGGVDNL FDKRLWRAGN AQT TGDLAGA 720
35 NYIAGAGAYT YNEPGRTWYM SVNTHF 746
<212> Type : PRT
<211> Length : 746
SequenceName : SEQ ID 57
SequenceDescription :

40 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
45 MGRFSLRYK KLSYRFVFLT LAGCSSVGNO SLKNETQESV KTKIVKGKTT KQDVLASFGE 60
PDSRSLIDGE EQW SYTYMNS QSKATSFIPV VGLLAGGADS QTKSLTVSFK GEKVSTYIFN 120
AGTSNVKTGI F 131
<212> Type : PRT
<211> Length : 131
50 SequenceName : SEQ ID 58
SequenceDescription :

Sequence

55 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MKKIACLSAL AAVLAFTAGT SVAATSTVTG GYAQSDAQGO MNKMGGFNLK YRYEEDNSPL 60
GVIGSFTYTE KSRTASSGDY NKNQYYGITA GPAYRINDWA SIYGVVGVGY GKFTTEYPT 120
YKHDTSDYGF SYGAGLQFNP MENVALDFS Y EQSRIRSV DV GTW IAGVGYR F 171
60
<212> Type : PRT
<211> Length : 171
SequenceName : SEQ ID 59
SequenceDescription :

65 Sequence

```
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MKSIALTVVC AISGIACVNL SAHAAEGEHT ISLGYAHFQF PGLKDFVKDA TAHNRETFSH      60
FVNRNYFSSL GEYTDGRVSG YEGKDKNPQG INIRYRYEIT DDFGVITSFT WTRSLTNSQT      120
5 FIDVQSADHT RRIKNPAASA RTDIRANYWS LLAGPSWRVN QYMSLYAMAG MGVAKVSADL      180
KIKDNINSSG GFSESNSTKK TSLAWAAGAQ FNLNESVTL D VAYEGSGSGD WRTSGVTAGI      240
GLKF      244
<212> Type : PRT
<211> Length : 244
10 SequenceName : SEQ ID 60
SequenceDescription :

Sequence
-----
15 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MRKLYAAILL AAICLTVSGA PAWASEQQAT LSAGYLHVST NAPGSDNLNG INVKYRYEFT      60
DTLGLVTSFS YAGDRNRQIT RYSDTRWHED SVNRNRWFSVM AGPSVRVNEW FSAYAMAGVA      120
20 YSRVSTFSGD YLRVTDNKGK THDVLTSDD GRHSNTSLAW GAGVQFNPTE SVAIDIAIEG      180
SGSGDWRTDG FIVGVGYKF      199
<212> Type : PRT
<211> Length : 199
SequenceName : SEQ ID 61
SequenceDescription :

Sequence
-----
25 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
30 MRKLYAAILL AAICLAVSGA PAWASEQQAT LSAGYLHART SAPGSDNLNG INVKYRYEFT      60
DTLGLVTSFS YAGDKNRQLT RYSDTRWHED SVNRNRWFSVM AGPSVRVNEW FSAYAMAGVA      120
YSRVSTFSGD YLRVTDNKGK THDVLTSDD GRHSNTSLAW GAGVQFNPTE SVAIDIAIEG      180
SGSGDWRTDG FIVGVGYKF      199
<212> Type : PRT
35 <211> Length : 199
SequenceName : SEQ ID 62
SequenceDescription :

Sequence
-----
40 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MRKLYAAILL AAICLAVSGA PAWASEQQAT LSAGYLHART SAPGSDNLNG INVKYRYEFT      60
DTLGLVTSFS YAGDKNRQLT RYSDTRWHED SVNRNRWFSVM AGPSVRVNEW FSAYAMAGVA      120
45 YSRVSTFSGD YLRVTDNKGK THDVLTSDD GRHSNTSLAW GAGVQFNPTE SVAIDIAIEG      180
SGSGDWRTDG FIVGVGYKF      199
<212> Type : PRT
<211> Length : 199
SequenceName : SEQ ID 63
50 SequenceDescription :

Sequence
-----
55 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MRKLYAAILL AAICLAVSGA PAWASEQQAT LSAGYLHART SAPGSDNLNG INVKYRYEFT      60
DTLGLVTSFS YAGDKNRQLT RYSDTRWHED SVNRNRWFSVM AGPSVRVNEW FSAYAMAGVA      120
45 YSRVSTFSGD YLRVTDNKGK THDVLTSDD GRHSNTSLAW GAGVQFNPTE SVAIDIAIEG      180
SGSGDWRTDG FIVGVGYKF      199
<212> Type : PRT
60 <211> Length : 199
SequenceName : SEQ ID 64
SequenceDescription :

Sequence
-----
65 <213> OrganismName : Escherichia coli O157:H7
```

<400> PreSequenceString :
MVMSQKTLFT KSALAVAVAI ISTQAWSAGF QLNEFSSSSGL GRAYSSEGAI ADDAGNVSRN 60
PALITMFDPR TFSAGAVYID PDVNISGTSP SGRSLKADNI APTAWVPNMH FVAPINDQFG 120
WGASITSNYG LATEFNDTYA GGSVGGTTDL ETMNLNLSGA YRLNNAWSFG LGFNAVYARA 180
5 KIERFAGDLG QLVAGQIMQS PAGKTPQGQA LAATANGIDS NTKIAHLNGN QWGFQWNAGI 240
LYELDKNNRY ALTYSSEVKI DFKGNYSSDL NRVFNNGYGL IPTATGGATQ SGYLTNLNLE 300
MWEVSGYNRV DPQWAIHYSY AYTSWSQFQQ LKATSTSGDT LFQKHEGFKD AYRIALGTTY 360
YYDDNWTFRF GIAFDDSPVP AQNRSISIPD QDRFWLSAGT TYAFNKDASV DVGVSVMHGG 420
SVKINEGPYQ FESEKAWLF GTNFNAYF 448

10 <212> Type : PRT
<211> Length : 448
SequenceName : SEQ ID 65
SequenceDescription :

15 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MAFSQAVSGL NAAATNLDVI GNNIANSATY GFKSGTASFA DMFAGSKVGL GVKVAGITQD 60
20 FTDGTTTNTG RGLDVAISQN GFFRLVDSNG SVFYSRNGQF KLDENRNLVN MQGLQLTGYP 120
ATGTPPTIQQ GANPTNISIP NTLMAAKTTT TASMQINLNS SDPLPSVNAF DASNADSYNK 180
KGSVTVFDSQ GNAHDMSVYF VKTGDNNWQV YTQDSSDPTG TAEPAMKLVF NANGVLTSNP 240
TENITTGAIN GAEPATFSL S FLNSMQQNTG ANNIVATTQN GYKPGDLVSY QINDDGTVVG 300
NYSNEQTQLL GQIVLANFAN NEGLASEGDN VWSATQSSGV ALLGTAGTGN FGTLTNGALE 360
25 ASNVDLSKEL VNMIVAQRNY QSNAQTIKTQ DQILNTLVNL R 401

<212> Type : PRT
<211> Length : 401
SequenceName : SEQ ID 66
SequenceDescription :

30 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
35 MSKSTFLHIL ISSIILVALI QSSAWANCTN TQIGQTEDGR TALIEFGKIN MTDITYFAPAG 60
SLLATTVVPP TNYTSGGATG SSVLWECDAT DLPNIYFLVA TNGDDRVGGF YDAGGPDGLS 120
DVYATWFAFV GLKQTMAGVT LGRYWKVPI TSYATQGTKI QIRLQDIPPL HAELYRISTL 180
PDTSATTSWC GNNNTDSSGV GFAKPSGTIY NCVQPNAYIQ LSGTSGILFG HDEPGEDSSV 240
HWDFWGADNG FGYGMRANR LYNNATCVAR SATPLVLLPT IAEAQLNAGM ESTGNFNVRV 300
40 ECSNSVQSGI SDTQTALGIQ VSEGAYTAAQ KLGIIINSNGG VSALVSDNYD AAEMAKGVGI 360
YISNSAHPDT AMLTVGQPGI AKLTPGGNAA GWYPVFEGAT LEGATHPGYS SYSYSFIARL 420
KKLPNQTVSA GKVRATAYIL VKMQ 444

<212> Type : PRT
<211> Length : 444
SequenceName : SEQ ID 67
SequenceDescription :

45 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MENNRRNFPAR QFHSLTFFAG LCIGITPVAQ ALAAEGQTNA DDTLVVEAST PSLYAPQOSA 60
DPKFSRPVAD TTRTMTVISE QVIKQGATN LTDALKNVPG VGAF FAGENG NSTTGDAIYM 120
RGADTSNSIY IDGIRDIGSV SRDTFNTEQV EVIKGPSGTD YGRSAPTGSI NMISKQPRND 180
55 SGIDASASIG SAWFRRTLD VNQVIGDTTA VRLNVMGEEK HDAGRDKVKN ERYGVAPSLA 240
FGLGTANRLY LNYLHVTQHN TPDGGIPTIG LPGYSAPSAG TATLNHSGKV DTHNFYGTDS 300
DYDDSTTDTA TMRFEHDIND NTTIRNTTRW SRVKQDYLMT AIMGCASNIT QPTSDVNSWT 360
WSRTANTKDV SNKILTNQTN LTSTFYTASI GHVSTGVEF TRETQTNNGV NPVTLPVNI 420
YHPDSSIHPG GLTRNGANAN GQTDTFIYA FDTLQITRDF ELNGGIRLDN YHTEYDSATA 480
60 CGGSGRGAIT CPAGVAKGSP VTTVD TAKSG NLVNWKAGAL YHLTENGNVY INYAVSQQPP 540
GGNNFALAQS GSGNSANRTD FKPQKANTSE IGTKWQVLDK RLLLTAAALFR TDIENEVEQN 600
DDGTYSQYQK KRVEGYEISV AGNITPAWQV IGGYTQKAT IKNGKDVAQD GSSSLPYTPE 660
HAFTLWSQYQ ATDDISVGAG ARYIGSMHKG SDGAVGTPAF TEGYWVADAK LGRVNRNLD 720
FQLNVYNLFD TDYVASINKS GYRYHPGEPR TFLLTANMHF 760

65 <212> Type : PRT
<211> Length : 760
SequenceName : SEQ ID 68

SequenceDescription :

Sequence

5 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MQMKLLLPIL IGLSLSGFSS LSQAENLMQV YQCARLSNPE LRKSAADRDA AFEKINEARS 60
PLLPQLGLGA DYTYSNGYRD ANGINSNATS ASLQLTQSIF DMSKWRALTL QEKAAGIQDV 120
TYQTDQQTLL LNTATAYFNV LNAIDVLSYT QAQKEAIYRQ LDQTTQRFNV GLVAITDVQN 180
10 ARAQYDITVLA NEVTARNNLD NAVEQLRQIT GNYYPELAAL NVENFKTDKP QPVNALLKEA 240
EKRNLSELLQA RLSQDLAREQ IRQAQDGHLP TLDLTASSGI SDTSYSGSKT RGAAGTQYDD 300
SNMGQNKVGL SFSLPYQGG MVNSQVKQAQ YNFVGASEQL ESAHRSVVQT VRSSFNNINA 360
SISSINAYKQ AVVSAQSSLD AMEAGYSVGT RTIVDVLDAT TTLYNKQEL ANARYNYLIN 420
QLNIKSALGT LNEQDLLALN NALSKPVSTN PENVAPQTPQ QNAIADGYAP DSEAPVWQQT 480
15 SARTTTSNHG NPFRRN 495
<212> Type : PRT
<211> Length : 495
SequenceName : SEQ ID 69
SequenceDescription :
20 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
25 MTKLKLALLG VLIATSAGVA HAEGKFSLGA GVGVEHPYK DYDQDVYPVP VINYEGDNFW 60
FRGLGGGYLL WNDATDKLSI TAYWSPLYFK AKDSGDHQMRL HLDNRKSTMM AGLSYAHFTQ 120
YGYLRRTTLG DTLDNSNGIV WDMAWLYRYT NGGLTVTPGI GVQWNSNQY EYYYGVSRKE 180
SARSGLRGYN SNDSWSPYLE LSASYNFLGD WSVYGTARYT RLSDEVTDSP IVDKSWTGLI 240
STGITYKF 248
30 <212> Type : PRT
<211> Length : 248
SequenceName : SEQ ID 70
SequenceDescription :
35 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
40 MKKTLAAGA VLALSSSFTV NAAENDKPKQV LSDWWHQSVN VVGSYHTRFG PQIRNDTYLE 60
YEAFKKDWF DFYGYADAPV FFGGNSDAKG IWNHGSPLFM EIEPRFSIDK LTNTDLSFGP 120
FKEWYFANNY IYDMGRNKDG RQSTWYMLG TDIDTGLPMS LSMNVYAKYQ WQNYGAANEN 180
EWDGYRFBK YFVPTDLWG GQLSYIGFTN FDWGSDDLGD SGNAINGIKT RTNNSIASSH 240
ILALNYDHHW YSVVARYWHD GGQWNDDAEL NFGNGNFNVR STGWGGYLVV GYNF 294
45 <212> Type : PRT
<211> Length : 294
SequenceName : SEQ ID 71
SequenceDescription :
50 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
55 MLSTQFNDRD QYQAITKPSL LAGCIALALL PSAAFAAPAT EETVIVEGSA TAPDDGENDY 60
SVTSTSAGTK MQMTQRDIPQ SVTIVSQORM EDQQLQTLGE VMENTLGISK SQADSDRALY 120
YSRGFQIDNY MVDGIPTYFE SRWNLGDAIS DMALFERVEV VRGATGLMTG TGNPSAAINM 180
VRKHATSREF KGDVSAEYGS WNKERYVADL QSPLTEDGKI RARIVGGYQN NDSWLDYNS 240
EKTFFSGIVD ADLGLDITLS AGYEQRIDV NSPTWGGLPR WNTDGSSNSY DRARSTAPDW 300
AYNDKEINKV FMTLKQRFAD TWQATLNATH SEVEFDSKMM YVDAYVNKAD GMLVGPYSNY 360
60 GPGFDYVGGT GWNSGKRKVD ALDLFADGSY ELFGRRHNLN FGGSYSKQNN RYFSSWANIF 420
PDEIGSFYNF NGNFPQTDWS PQSLAQDDTT HMKSLYAATR VTLADPLHLI LGARYTNWRV 480
DTLTYSMEKN HTTPYAGLVF DINDNWSTYA SYTSIFQPQN DRDSSGKYLA PITGNNYELG 540
LKSDDWMSRL TTTLAIFRIE QDNVAQSTGT PIPGSNGETA YKAVDGTVSK GVEFELNGAI 600
TDNWQLTFGA TRYIAEDNEG NAVNPNLPRT TVKMFTSYRL PVMPELTVGG GVNWQNRVYT 660
65 DTVTPYGTFR AEQGSYALVD LFTRYQVTKN FSLQGNVNNL FDKTYDTNVE GSIVYGAPRN 720
FSITGTQF 729
<212> Type : PRT

<211> Length : 729
SequenceName : SEQ ID 72
SequenceDescription :

5 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MARFQFKNRK NNGLIFFISF MVMGEAAIAA PLPQWANAPA VTPVAQLSLQ ESILRAFARN 60
10 PGVTTQAAQI GIGEAQIDEA KSAWYPHVGL TGNAGPSRQT DSSGRLDNNV SYGITLTQLV 120
YDFGKTNDI NLQTAARDSY RFKLMATLTD VAEKTATAYM EVSRYQALCD AAQRNIHSL 180
NVYNMAALRA NAGLNSSSDE LQAQTRIAGM RSTLEQYQAA MASAKAQLAV LTGVQPEAIA 240
APPAELAEQP VSLKNIDYQS IPLVLAAENL RQSAQYGVK TKAQYWPTLS IQGGKTRYQT 300
SDRSYWDDQL QLVNVPAPLYQ GGAVSAQVQQ AEGQQKISAS QVEQAKLDVL QRASVAYANW 360
15 TGARGREEAG LAQSESAHKT RDVYQNEYKL GKRLNDLLT VEQDVFAQS AEINANYDGW 420
VAAVNYAAAV NNLIPLAGIK QGLYNDLPDL K 451
<212> Type : PRT
<211> Length : 451
SequenceName : SEQ ID 73
20 SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MAKFTSPFSG IKGRALFSLI FAAPMIHATD TATTKDGETI TVTADANTAT EATDGYQPLS 60
TSTATLTDMPLMDIPQVNT VSDQVLENQN ATTIDEALYN VSNVQNTNL GGTQDAFVRR 120
GFGANRDGSI MTNGLRTVLP RSFNAATERV EVLKGPASTL YGILDPGGLI NVVTKRPEKT 180
FHGSVSATSS SFGGGTGQLD ITGPIEGTQL AYRLTGEVQD EDYWRNFGKE RSTFIAPSLT 240
30 WFGDNATVTM LYSHRDYKTP FDRGTIFDLT TKQPVNVDRK IRFDEPFNIT DGQSDLAQLN 300
AEYHLNSQWT ARFDYSYSQD KYSDNQARVT AYDATTGTLT RRV DATQGST QRMHSTRADL 360
QGNVDIAGFY NEILGGVSYE YYDLLRTDMI RCKNAKDFNI YNPVYGNITSK CTTVSASDSD 420
QTIKQESYSA YAQDALYLTDL NWIAVAGIRY QYYTQYAGKG RPFNVNTDSR DEQWTPKLGL 480
VYKLTSPSVSL FANYSQTFMP QSSIASYIGD LPPESSNAYE VGAKFELEFDG ITADIALFDI 540
35 HKRNVLYTES IGETIAKTA GRVRSRGVEV DLAGALTENI NIIASVGYTD AKVLEDPDYA 600
GKPLPNVPRH TGSFLTYDI HNMPPGNNTLT FGGGHCVSR RSATNGADYY LPGYFVADAF 660
AAYKMKLQYP VTLQLNVKNL FDKTYTSSI ATNNLGNQIG DPREVQFTVK MEF 713
<212> Type : PRT
40 <211> Length : 713
SequenceName : SEQ ID 74
SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MRTLQGWLLP VFMLPMAVYA QEATVKEVHD APAVRGSIIA NMLQEHDPNF TLYPYDTNYL 60
IYTQTSIDLNK EAIASDWAE NARKDEVKFQ LSLAFPLWRG ILGPNVSLGA SYTQKSWWQL 120
50 SNSEESSPFR ETNYEPQLFL GFATDYRFAG WTLRDVEMGY NHDSNGRSDP TSSRWNRLYT 180
RLMAENGWNL VEVKPYVVG NTDDNPDIK YMGYYQLKIG YHLGDAVLSA KGQYNWNTGY 240
GGAELGLSYP ITHKVRLYTQ VYSGYGESLI DYNFNQTRVG VGVMLNDLF 289
<212> Type : PRT
<211> Length : 289
55 SequenceName : SEQ ID 75
SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MAVQKNVIK ILAGTFALML SGCVTVPDAI KGSSPTPQQD LVRVMSAPQL YVGQEARFGG 60
KVVAVQNQQG KTRLEIATVP LDGARPTLG EPSRGRIYAD VNGFLDPVDF RGQLVTVVGP 120
ITGAVDGKIG NTPYKFMVMQ ATGYKRWHLT QQVIMPPQPI DPWFYGGRGW PYGHGGGWGY 180
65 NPGPARVQTV VTE 193
<212> Type : PRT
<211> Length : 193

SequenceName : SEQ ID 76
SequenceDescription :

5 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MRKQWLGICI AAGMLAActs DDGQQQT VSV PQPAVCNGPI VEISGADPRF EPLNATANQD 60
YQRDGKSYKI VQDPSRFIQA GLAAIYDAEP GSNLTASGEA FDPTQLTAAH PTLPIPSYAR 120
10 ITNLANGRMI VVRINDRGPY GNDRVISLSR AAADRLNTSN NTKVRIDPII VAQDGSLSGP 180
GMACTTVAKQ TYALPAPPDL SGGAGTSSVS GPQGDILPVS NSTLKSEDPT GAPVTSSGFL 240
GAPTTLAPGV LEGSEPTPAP QPVVTAPSTT PATSPAMVTP QAASQSASGN FMVQVGAVSD 300
QARAQQYQQQ LGQKFGVPGR VTONGAVWRI QLGPFANKAE ASTLQQRLQT EAQLQSFITT 360
AQ 362

15 <212> Type : PRT
<211> Length : 362
SequenceName : SEQ ID 77
SequenceDescription :

20 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MIKRVLVISM VGLSLVGCVN NDTLSGDVYT ASEAKQVQNV SYGTIVNVRP VQIQGGDDSN 60
25 VIGAIGGAVL GGFLGNTVGG GTGRSLATAA GAVAGGVAGQ GVQSAMNKTQ GVELEIRKDD 120
GNTIMVVQKQ GNTRFSPGQR VVLASNGSQV TVSPR 155
<212> Type : PRT
<211> Length : 155

30 SequenceName : SEQ ID 78
SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
35 <400> PreSequenceString :
MSKATEQN DK LKRAIIISAV LHVILFAALI WSSF DENIEA SAGGGGGSSI DAVMVD SGAV 60
VEQYKRMQS Q ESSAKRSDEQ RKMKEQQAAB ELREKQAABEQ ERLKQLEKER LAAQE QKKQA 120
EEAAKQAE LK QKQAE EAAAK AAADAKAKAE ADDKAAEEAA KKAADAKKK AEAEAAKAAA 180
EAQKKA EAAA AALKKKA EAA EAAAAEARKK AA AEKAAADK KAAEKAAAEK AAADK KAAAE 240
40 KAAADK KAAA AKAAAEKAAA AKAAAEADDI FGELSSGKNA PKTGGGAKGN NASPAGSGNT 300
KNNGASGADI NNYAGQIKSA IESKFYDASS YAGKTCTLRI KLAPDGM LLD IKPEGGDPAL 360
CQAALAAAKL AKIPKPPSQA VYEVFKNAPL DFKP 394

<212> Type : PRT
<211> Length : 394
45 SequenceName : SEQ ID 79
SequenceDescription :

Sequence

50 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MMKF KKCLLP VAMLASFTLA GCQSNADDHA ADVYQTDQLN TKQETKTVNI ISILPAKVAV 60
DNSQNKRNAQ AFGAMNIGAVA GGVIGHNVGS GSNSGTTAGA VGGGAVGAAA GSMVNDKTLV 120
EGVSLTYKEG TKVYTSTQVG KECQFTTGLA VVITTTYN ET RIQPNTKCPE KS 172

55 <212> Type : PRT
<211> Length : 172
SequenceName : SEQ ID 80
SequenceDescription :

60 Sequence

<213> OrganismName : Escherichia coli O157:H7
65 <400> PreSequenceString :
MLLSIITVAF RNLEGIVKTH ASLAHLAQAE DISFEWIVVD GGSNDGTREY LENLNGIYNL 60
RFSVSEPDNGI YDAMNKGIAM AQKFALFLN SGDIFHQDAA YFVRKLKMQK DNVMITGDAL 120
LDFGDGHKIK RSAKPGWYIY HSLPASHQAI FFPVSGLKKW RYDLEYKVSS DYALAAKMYK 180

AGYAFKKLNG LVSEFSMGGV STTNMELCA DAKKVQRQIL HVPGFWAELS WHLRQRTTSK 240
TKALYNKS 248
<212> Type : PRT
<211> Length : 248
5 SequenceName : SEQ ID 81
 SequenceDescription :

Sequence

10 <213> OrganismName : Haemophilus influenzae Rd
 <400> PreSequenceString :
 MKLTTLQTLK KGFTLIELMI VIAIIAILAT IAIPSYQNYT KKAASSELLQ ASAPYKADVE 60
 LCVYSTNETT SCTGGKNGIA ADIKTAGGYV ASVITQSGGI TVKNGGTLAN MEYILQAKGN 120
 AAAGVTWTTT CKGTDASLFP ANFCGSVTK 149
15 <212> Type : PRT
 <211> Length : 149
 SequenceName : SEQ ID 82
 SequenceDescription :

20 Sequence

 <213> OrganismName : Haemophilus influenzae Rd
 <400> PreSequenceString :
25 MLNKKFKLNF IALTVAAYALT PYTEAALVRD DVDYQIFRDF AENKGRFSVG ATNVEVRDKN 60
 NHSLGNVLPN GIPMIDFSV DDKRIATLI NPQYVVGKHH VSNQVSELHF GNLNGNMNNG 120
 NAKSHRDVSS EENRYFSVEK NEYPTKLNGK AVTTEDQTQK RREDYYMPRL DKFVTEVAPI 180
 EASTASSDAG TYNDQNKYPA FVRLGSGSQF IYKKGDNYSL ILNNHEVGGN NLKLVGDAYT 240
 YGIAGTPYKV NHENGLIGF GNSKEEHSDF KGILSQDPLT NYAVLGDSGS PLFVYDREKG 300
 KWLFLGSYDF WAGYNKKSQW EWNIIKPEFA KTVLDKDTAG SLTGSNTQYN WNPTGKTSVI 360
30 SNGSESLNVD LFDSSQDSDS KKNHNGKSVT LRSGTLTLN NNIDQAGGL FFEGDYEVKG 420
 TSDSTTWKGA GVSVDGKTV TWKVHNPNSD RLAKIGKGTI IVEGKGENKG SLKVGDTVI 480
 LKQADANNK VKAFSQVGIV SGRSTVVLND DKQVDPNSIY FGFRGGRDLA NGNNLTFEHI 540
 RNIDDGARLV NHNTSKTSTV TITGESLITD PNTITPYNID APDEDNPYAF RRIKDGGLY 600
 LNLNENYTYA LRGGASTRSE LPKNSGESNE NWLYMGKTSI EAKRNVNMHI NNERMNGFNG 660
35 YFGEEGKNN GNLNVTFK GK SEQNRFLLTG GTNLNGDLKV EKGTLFLSGR PTPHARDIAG 720
 ISSTKKDQHF AENNEVVVED DWINRNFKAT NINVTNNATL YSGRNVANIT SNITASDNAK 780
 VHIGYKAGDT VCVRSDYTG YCTTDLKSD KALNSFNATN VSGNVNLSGN ANFVLKANL 840
 FGTISGTGNS QVRLTNSHW HLTGDSNVNQ LNLDKGHIHL NAQNDANKVT TYNTLTVNSL 900
 SGNGSFYILT DLSNQGGDKV VVTKSATGNF TLQVADKTGE PTKNELTLFD ASNATRNMLN 960
40 VSLVGNVTDL GAWKYKLRNV NGRYDLYNPE VEKRNQTVDT TNITTPNNIQ ADVPSVPSNN 1020
 EELIARVETPV PPPAPATPSE TTETVAENSK QESKTVEKNE QDATETTAQN GEVAEEAKPS 1080
 VKANTQTNEV AQSGSETEET QTTEIKETAK VEKEEKAKVE KDEIQEAPQM ASETSPKQAK 1140
 PAPKEVSTDT KVEETQVQAQ PQTQSTTVAA AEATSPNSKP AEETQPSSEKT NAEVTPPVVS 1200
 KNQNTENTDQ PTEREKTAKV ETEKTQEPPO VASQASPKQE QSETVQPPAV LESENVPVNV 1260
45 NAEEVQAQLQ TQTSATVSTK QPAPENSINT GSATAITETA EKSDKPKQET AASTEDASQH 1320
 KANTVADNSV ANNSSESDPK SRRRRSISQP QETSAEETTA ASTDETTIAD NSKRSPKPNR 1380
 SRRSVRSEPT VTNGSDRSTV ALRDLTSTNT NAVISDAMAK AQFVALNVGK AVSQHISQLE 1440
 MNNEGQYNVW VSNTSMNENY SSSQYRRFSS KSTQTQLGWD QTISNNVQLG GVFTYVRNSN 1500
 NFDKASSKNT LAQVNFYSKY YADNHWYLG DLGYGKFQSN LKTNHNAKFA RHTAQFGLTA 1560
50 GKAFNLGNFG ITPIVGVRY YLSNANFALA KDRIKVPIS VKTAFQAQVD SYTYHLGEFS 1620
 VTPILSARYD TNQSGSKINV NQYDFAYNVE NQQQYNAGLK LKYHNVKLSL IGGLTKAKQA 1680
 EKQKTAELKL SFSF 1694
 <212> Type : PRT
 <211> Length : 1694
55 SequenceName : SEQ ID 83
 SequenceDescription :

Sequence

60 <213> OrganismName : Haemophilus influenzae Rd
 <400> PreSequenceString :
 MALVNKIKTL SSVGILAAATL FLAGCQAQSN ILAFTPPAPS ASMNVNRTAV VSVTTKDSRA 60
 IQEIASYTKH GELIKLNASP SVTQLFQQVM QQNLISSKGF VGQLNGSNAW VTVDVREFGT 120
 QVEQGNLRYK LNTKIQTAVY VQGAQGSYNK SFNVTHSQEG VFNAGNDEIH KVLSTQTFNDI 180
65 VNNIYQDQEV AAAINQYSN 199
 <212> Type : PRT
 <211> Length : 199

SequenceName : SEQ ID 84
SequenceDescription :

Sequence

5 -----
<213> OrganismName : Haemophilus influenzae Rd
<400> PreSequenceString :
MLCWIGYKNG ILPQQNSTLY PWNPSKCGV IFDGFQLVGD DFNSDQTAEN TSPAWQVLYT 60
THLQSCSPIH SGENFAPIPL YKQLKNQPHL SQDLIKWQEN WQACDQLQMN GAVLEQQSLA 120
10 EISDHQSTLS KHGRYLAQEI EKETGIPTY YLYRVGGQSL ESEKSRCCPS CGANWALKDA 180
IFDTFHFKCD TCRLVSNLSW NFL 203
<212> Type : PRT
<211> Length : 203
SequenceName : SEQ ID 85
15 SequenceDescription :

Sequence

<213> OrganismName : Haemophilus influenzae Rd
20 <400> PreSequenceString :
MGAFAFASVT NANIYAEGDI GLSQTKANGS NNTRVGPRVS VGYKVGNTRV AGDYTHHGKV 60
DGTKIQGLGA SVLYDFDFTNS KVQPYVGARV ATNQFKYTNR AEQKFKSSSD IKLGYGVVAG 120
AKYKLDGNWY ANGGVEYNRL GNFDSTKVN YGAKVGVGYG F 161
<212> Type : PRT
25 <211> Length : 161
SequenceName : SEQ ID 86
SequenceDescription :

Sequence

30 -----
<213> OrganismName : Haemophilus influenzae Rd
<400> PreSequenceString :
MKKLLIASLL FGTTTTFVFAA PFVAKDIRVD GVQGDLEQQI RASLPVRAGQ RVTDNDVANI 60
VRSLFVSGRF DDVKAHQEGD VLVVSVVAKS IISDVKIKGN SIIPTEALKQ NLDANGFKVG 120
35 DVLIREKLNE FAKSVKEHYA SVGRYNATVE PIVNTLPNNR AEILIQINED DKAKLASLTF 180
KGNESVSSST LQEQMELQPD SWWKLWGNKF EGAQFEKDLQ SIRDYYLNNG YAKAQITKTD 240
VQLNDEKTKV NVTIDVNEGL QYDLRSARII GNLGMSAEL EPLLSALHLN DTFRRSDIAD 300
VENAIKAKLG ERGYGSATVN SVPDFDDANK TLAITLVVDA GRRLTVRQLR FEGNTVSADS 360
TLRQEMRQQE GTWYNSQLVE LGKIRLDRTG FFETVENRID PINGSNDEV VVYKVKERNT 420
40 GSINFGIGYG TESGISYQAS VKQDNFLGTG AAVSIAGTKN DYGTSVNLGY TEPYFTKDG V 480
SLGGNVFFEN YDNSKSDTSS NYKRTTYGSN VTLGFPVNE NSYYVGLGHT YNKISNFALE 540
YNRNLYIQSM KFKGNGIKTN DDFDSFGWNY NSLNRGYFPT KGVKASLGGR VTIPGSDNKY 600
YKLSADVQCF YPLDRDHLWV VSAKASAGYA NGFGNKRLPF YQTYTAGGIG SLRGFAYGSI 660
GPNAIYAEHG NGNGTFKKIS SDVIGGNAIT TASAELIVPT PFVSDKSQNT VRTSLFVDAA 720
45 SVWNTKWKSD KSGLDNNVLK SLPDYGKSSR IRASTGVGFQ WQSPIGPLVF SYAKPIKKYE 780
NDDVEQFQFS IGGSF 795
<212> Type : PRT
<211> Length : 795
SequenceName : SEQ ID 87
50 SequenceDescription :

Sequence

<213> OrganismName : Haemophilus influenzae Rd
55 <400> PreSequenceString :
MLKKTSLIFT ALLMTGCVQN ANVTTPQAQK MQVEKVDKAL QKGEADRYLC QDDRVRVVRVH 60
ATHKKYKKNL HYVTVTFGV SEKLTLMISE RGKNYANIRW MWQERDDFST LKTNLGEBILA 120
TQCVSQTSE LSGQ 134
<212> Type : PRT
60 <211> Length : 134
SequenceName : SEQ ID 88
SequenceDescription :

Sequence

65 -----
<213> OrganismName : Haemophilus influenzae Rd
<400> PreSequenceString :

MRIIIIFFMG LNMNTNFRLER ACLFRYAWAN GRCCCLCSSTN OPTNQPTNQPTN TNQPTNQPTN 60
QPTNQNSNVSEQLEQINVSG STENS DTKTP PKIAETVKTA KTLEREQANN IKDIVKYETG 120
VTVVVEAGREFG QSGFAIRGVD ENRVAINIDG LRQAETLSSQ GFKELFEGYG NFNNTNRNGAE 180
IETLKEVNIT KGADSIKNGS GSLGGSVIYK TKDARDYLIN KDYYVS YKKG YATENNQSFD 240
5 TLTLAGRYKK FDDLIVVTTSR NGHELENYGY KNYNDKIQGG KREKADPYKI EQDSTLLKLS 300
FNPTENHRFT FAADLYEHRS RGQDLSYTLK YQRCNETPE VDSRHTNDKT KRRNISFSYE 360
NFSQTPFWD LKLTYSQDRI KTRARTDEYC DAGVRHCEGT DNPTGLKVTN GKITRRDGS 420
LQFEEKNNTA KSSDKTYDFK KPIDTDKRV DDKLVNLNPS DTWYDCSIFN CENNAKIKVF 480
KGNYYGYDG KWKVDLEIK ELNGKKFAKI KDNDRIKISI LPSSPGYLER LWQERDLDTN 540
10 TQQLNLDLTK DFKIWHIEHN LQYGGSYNTA MKRMVNRAGN DASDVQWWAT PTLGEDSWTG 600
KPHTCATTYE WNANLCPRVD PEFSYLLPIK TTGKSVYLFDFNFVITDYLSDLG YRYDNIH 660
YQPKYKHGIT PKLPDDIVKG LFIPLPNNSN SDPNKVKENV QONIDYIAKQ NKKYKAHSYS 720
FVSTIDPTSF LRLQLKYSKG FRTPTSDEMY FTFKHPDFTI LPNTDLKPEI AKTKETIAFTL 780
HNDDWGFIST SLFKTMYKMF IDLIFKKQET FVGGSGRGE TLFFSLYQNI NRDNASL KGI 840
15 EINSKVFLGK MAKPMDFNL SYKYTYQGR MNGNIPMNAI QPRTMVYGLG YDHPNHKFGF 900
DFYTHVASK NPEDTYNMFY KEENKKDSTI KWRKSYTIL DLIGYVQPIK NLTRAGVYN 960
LTNRKYYTWD SRSRYSFGT SNVIDQSTGL GINRFYAPGR NYKMSVQPEF 1010
<212> Type : PRT
<211> Length : 1010
20 SequenceName : SEQ ID 89
SequenceDescription :

Sequence

25 <213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MTYRNGKIDL KERFSKNRSF KGIKKKIACK YTIKNSLSII YSLKTHSNSS LSINKKIFLG 60
LGFVSALSAQ SEDYNSSVYW LNSVNNNNNN KSYIISPLRT WAGGNRSFTQ NNNNSQLYIG 120
TKNASATPNH SSVWFGEKGY IGFITGVFKA RDIFITGAVG SGNELKTGGG AILVFESSNE 180
30 LTTNGAYFQN NRAGTQTSWI NLISNNSVNL TNDTDFGNQTP NGGFNVMG RK ITYNGGSSVNG 240
GNFGFDNVDS NGATTISGVT FNNNGALTYK GGNGTGGSI FTNSNINHYK LNLNANSVTF 300
NNSTLGSMPN GNANTIGNAY ILNANNITFN NLTFNGGWV FNRSDAHVNF QGTTTINNPT 360
SPFVNMTGKV TINPNAIFNI QNYTPTIGNA YTLFSMKNGN IAYDDVNNLW NIIRLKNTQA 420
TKDNSKNATS NNNTHTYVYT YNLGGTLYHF RQIFSPDSIV LQSVYYGANN LYYTNSVNIH 480
35 DNVFNLKNIN DDRADTIFYL NGLNTWNTYQ ARFAQTYGGK NSALVFNAAT PWANGAIPKS 540
NSTVRFGGYE GVNWGKTGYI TGTFTADRVY ITGNMMSGNG AQTGGGATLN FVGATEINIA 600
GATFKNLKT SONSMTFMA LGNGSGSGKI NVSQSDFYDW TDGGYDFTGN GVFDVNFNK 660
AAYKFQGAEN SYNFKNTNFI AGNFKFQGKT TIEKSVLND SYAFDGVNNA FNEDKFNNGS 720
FNFNHAQTN AFNNSFSFGG SFSFNAKQVD FNGNSFNGGV FNFNNTPKAS FTNDTFNVNN 780
40 QFKINGAQT FTFSGKVVF MQGLLSSLSV GTTYQLLNK SVGYKDNNA LYQMLRWTSG 840
ENPSGKLVD NKTAPNSAKI YNVQFTDNGL TTYIKENFNN GITLTRLCTL GYTHCVNIDN 900
DAFNLKNVNN NASNTVFYLN GMTTWKTAGT GVFTQDYSGT NSVLVFNQTT PFLAGANPTS 960
NSVVGFGKTS GAEWGLVGYI QGVFKANQID ITGTIRSGNG AKTGGGATLV FNAQERLNTA 1020
NANLNNDKAG LQNSWMNFIV NNGNLNVINA NFSNQTPHGG FNLKANNITW DKGSVSGGGN 1080
45 FGVDNANANG NAVIKNVNFS DNGTLIYKGG ENSAGNSLTL ENNTFNSYNI NAKAQNLI FN 1140
NNSFNSSGYS FNDTKNVTFK GTNTLINSDF FSRLKGSVSI DNNSIFNIE DLTDKTTYTL 1200
LSGDNIKYNN QALADNVFSK NLWDLIHYDG EQGTLRLTDN NTYFVQFTQS NGQKFVFEET 1260
FNPGSITYKY FTIHSSPFHT EADSKDIWNQ VRKQDFDIPG KTPVCVGVCY IAPYKNQDLI 1320
GSSAFAWSLN FGATVVGTL LGSQAQKANN NGGSIWFGKN NLLYLHGNFN ATNIFLTNNF 1380
50 NVGNPNAGGG ATINFNADET LSADGLNYTN FQTVMAGLQT SASQHSWANF NSKLSMEIKN 1440
SNFRDFTWGG FRFNSGRITF ENTTFSGWTN INGATESGSS YVMNVANTDL IFTDSLGGG 1500
IRYDLKANNI IFNNTQMVD VSKNVNQSSL NGNVTFNHSR LSVKPNAAIN IGGDQTQTTL 1560
ENASSLSFYN DSVANFNGT AFNGVSYLNL NPNQVSVFN ANFNANVTF YGIPLFGKTP 1620
NFGNSVRLIN FKGDAKFNA TLNLRKNIH LNFQGAFTF NNSTMNLAES SQASFNALS 1680
55 EGETNFNLNG SLLSFNGNS VFNAPVNFYA NNSQISFTHS ATFNADASFD LGNNSTLNFQ 1740
SVLLNSALNL LGNGGNLAI NAKGNFSFGS QGILNLSYMN LFGGDKKASV YDVLQAQNI 1800
GLRGNGGYEK IRFYGIQIEK ADYSFNNGVH SWSFTNPLNT TETITETLHN NRLKVQISQN 1860
GASNNAMFNL APSLYDYQQN PYDESENSYN HTSDKAGTY LSSSIKGF GK NNEIPGTYNA 1920
QNQPLQALHI YNQAIKQDL NMIASLGKEF LPKVAKLIAS GALDNLNLNS PDSFETIFSI 1980
60 LKEYGITLNG ANWKSLLKII NNFNTANYH FSQGSVVGA IKEGQNTNS VVWFGGDGYK 2040
NPCAVGDNTC QMFRQTNLQ LLNSSVPYLG YINANFKAKN IYITGTIGSG NAWGSGGSAN 2100
VSFESATNLV LNQANIDAQG TDKIFSYL GK EGIDKLFGEK GLGNVLSNIV YVESLNDNAI 2160
PKDLANMIPK DLGSKTSLSS LSPTEVNLL GVSAPKNAIM EILNSKTVD VFGENGLLNA 2220
LDPVKRKEID QMLLEQIQAH SSGFEKFIVK TLGIENVENF INNWYKQSL SSFANNFVPG 2280
65 GLNQALDKIG SSSDAKDLQS FLDKTTFGDI LNQMINTQAPL INKLISWLGP QDLSVLVNI 2340
LNSITNPSKE LLGAISGMGQ KVLNDLLGEG VVNKIMSNQV LGQMINKIIA DKGFGGVYHQ 2400
GLGSILPKSL QDELKKLG MG SLLKPKGLHN LWQKGNFNFV AKNHVFNNS LFSNATGGEL 2460

	NFVAGKSIIF	NGKNTINF	TQ	YQGRLSFVSK	DFSNISLDTL	NATNGLTLNA	SKNDISVQKG	2520
	QICVNVLD	CM	TAKGKTTQ	TN	SSSSATAP	TN	ETLEVSAN	2580
	TTIGTLDL	G	NATPKANN	LI	VNNAFN	NNSN	YRANISGN	2640
	NSEGPLIF	N	LN	NP	TH	Q	TI	2700
5	NQTILG	SSLS	DY	LK	LY	TL	ID	2760
	SSILYDK	IQV	TV	SD	K	P	MS	2820
	NPLFAP	Y	LQ	DN	TE	H	IV	2880
	NFASFD	ST	D	SE	R	L	SS	2940
	TGTLYG	V	NV	G	D	Y	RV	3000
10	SVNETW	G	ANK	TQ	ISS	N	D	3060
	YYYIGM	S	G	L	E			3120
	LLVNSM	G	D	K	L			3180
	MIDIIGN	I	G	M				3194

<212> Type : PRT
 <211> Length : 3194
 SequenceName : SEQ ID 90
 SequenceDescription :

Sequence

 <213> OrganismName : Helicobacter pylori J99
 <400> PreSequenceString :

	MKQFKKKPKK	IKRSHQNQKT	ILKRPLWLMP	LLIGGFASGV	YADGTDILGL	SWGEKSQKVC	60
	VHRPWYAIWS	CDKWEKTTQ	FTGNQLITKT	WAGGNAANY	Y	HSQNNQDITA	120
25	LSGLYNYTGG	EYNGGNLDIE	LGSNATFNLG	ASSGNSFTSW	YPNGHTDVT	F	180
	VEVGNRVSG	AGTHTGTATL	NLNANKVTIN	SNISAYKTSQ	VNVGNANSVI	TINSVSLNGD	240
	TCSSLARVGV	GANCSTSGPS	YSFKGTTNAT	NTTFSNSSGS	F	TFEENATFS	300
	FNKKFNATNN	TAFNSGSFTF	KGTSSFNAN	FSNASYTFNN	Q	ATFQNSSFN	360
	NQSTQHPQIQ	NSSFSGSATT	LKGFATFEQA	FNNSNHQLTI	Q	NASFNNATF	420
30	DASFNNTSFN	TPVDTNMTI	SGGVTLGSKN	DLKNGATLDF	G	SSKITLTQ	480
	EKSVTILNSR	GGITYNHLLN	HAINSLTNAL	KTNESSSKPQ	S	FAQGLWDM	540
	NENAATSKPT	DSSPSKSSTN	STQVYQVGYK	IGDTIYKLQE	T	FSHNSIIIQ	600
	PVINGSKFDL	SASNYINADM	PWYNHKYYIP	KSNFTESGT	Y	YLPSPVQIW	660
	SASNSNLVIG	YNATWTDHNV	SSSDTVAFGD	TSGSALNGHC	G	PWPYYQCTG	720
35	VYITANLRSG	NRIGTGGAAN	LIFNGVDSIN	IANATTQH	N	AGAYSSSMTF	780
	LNGLNSNGKL	LVYGTFTTNQ	AKDGKFI	FNA	Q	ATFENTNF	840
	NSGSFETGAK	NTIFNNANFN	NSTSFNFNNS	SATTSFVGDF	T	NANSNLQIA	900
	GSQNTANFNN	TGSVNLAGNA	TFDNVVFNSP	TNTSVKGKVT	L	NNITLKNLN	960
	VFSAHSVINI	GEAITNGNPI	TLVSSSKAIE	YNDAFSKNLW	Q	LINYYQGHGA	1020
40	NGVYDVVYSE	NNQTYNFQEV	FSPNSISIRR	LGVGMVFDYV	D	MEKSDRLYY	1080
	PNSYNNNLGN	LNNITYYYDN	SIDFYASGKT	LFTKAEFSQT	F	TGQNSAIVF	1140
	DAPQSNVIIR	FGDNKGAGSN	DASGHWCWNLQ	CIGFITGHE	A	KIYITGSI	1200
	GASLNFNLGQ	GILLTNATLY	NRAAGTQSSS	MNFVNSANSI	Q	AQNSYFIDD	1260
	SFNALNLDFS	NSSFRGYVQG	TQSVFKFNAV	NAISFTNSSN	L	SSGLYQMQA	1320
45	SVSVGTSSIK	ANAINLSQNA	SINASNHSTL	ELQGDILNLD	T	SSLNLNQA	1380
	DYASLIASNG	SHLNFNGAVN	FNSANITTS	L	SSSSIVFKGA	VSLRGQFNLS	1440
	SAITSNTAFN	FYDNAFSQSP	ITFHQALDIK	VPLSLGGLNL	N	PNNSSVLNL	1500
	GSLNIANIDL	LSDLNGNKNR	VYNIQADMN	GNWYERINFF	G	MRINDGIYD	1560
	PLNNALKITE	SFKNNQLSVT	LSQIPGIKNT	LYNIGSEIFN	Y	QKVYNNANG	1620
50	VFYLTSSVK	G	Y	YNPNQSYQA	SGSNNTTKNN	NLTSESSVIS	1680
	NFSNIKALGQ	MALKLYPEIK	KILGNDFSL	SLSNLKGDAL	N	QLTKLITPS	1740
	NANNSVVQNF	NNGTLIIGAT	KIGQTDNTSA	VVFGGLGYQK	P	CDYTDIVCQ	1800
	LESISADLGY	INTTFNAKEI	YLTGTLGSGN	AWGTGGSASV	T	FNSQTSLLI	1860
	DGIFSMLGQE	GINKVFNQAG	LANILGEVAM	QSINKAGGLG	N	LIVNTLGSD	1920
55	QKNQTLSQLL	GQNNFNDLMN	DSGLNTAIKD	LIRQKLGFWT	G	LVGGLAGLG	1980
	IGSMSINDLL	SKKGLFNQIT	GFISANDIGQ	VISVMLQDIV	K	PSDALKNDV	2040
	FLGQDTLNSL	ESLLQNQQIK	SVLDKVLAAK	GLGSIYEQGL	G	DLIPNLGKK	2100
	VWQKGDFSFN	AQGNVVFQNS	TFSNANGGTL	SFNAGNSLIF	A	GNNHIAFTN	2160
	QVSNINVTML	NASNGLKINA	TNNNVSVSQG	NLFINASCVQ	Q	SDPTTASAT	2220
60	SSSNASNAP	IALNNNDESL	VVTANGFNFS	GNIYANGVVD	F	SKIKGSANV	2280
	FQANNLTISN	QAVLEKNASF	VTNNLIQGA	FNNNATQKIE	V	LQNLVIASN	2340
	EVGGALNNLG	AIHFNLNSQ	TPVNPLIQVG	GIINLNTTQT	P	FMNVSVANG	2400
	YIDYNINPNS	LQSYLKYTL	ININGNHIEE	KNGVLTYLQ	R	VLLQDKGLL	2460
	ASQNNILSLS	VLHNQIKMSY	GNKVMDFTPP	TLQDYIVGIQ	G	QSALNQIEA	2520
65	STLMMETKEN	PLFAPYILEN	HSLNEILGVT	KDLQNTASLI	S	NPNFRNNAT	2580
	QTSRLTKLSD	FRAREGESNF	SERLLELKNK	RFSDPNPSEV	F	VKYSQLSKH	2640
	GASFISGGNG	TLYGLNVGYD	RLVKSVILGG	YVAYGYSGFN	G	NIMHSLANN	2700

LKRNEFTLSA NETYGGNASH INSSNSLLSV LNQRNYNTW TTSVNGNYGY DFMFKQKSVV 2760
 LKPQVGLSYH FIGLSGMKGK MQNPAYQQFV MHSNPSNESV LTLNMGLESR KYFGKNSYYP 2820
 VTARLGRDLL IKAKGDNVVR FVGENTLLYR KGEIFNTFAS VITGGEMHLW RLMYVNAGVG 2880
 LKMGLQYQDL NITGNVGMVRV AF 2902

5 <212> Type : PRT
 <211> Length : 2902
 SequenceName : SEQ ID 91
 SequenceDescription :

10 Sequence

 <213> OrganismName : Helicobacter pylori J99
 <400> PreSequenceString :

15 MAFKKARLIS RFISKGSFKL NKISKKFFTL NQILKREKPI KRHKKTKSIE KPFNKKNKSFL 60
 KASVLLIGAL GGLSHLRANE CRYWSWSSWS YQDNIESGPN SPTHNSYCLF SSAQGSSTYY 120
 LNTLTYSAG GASFTQKFNG GTLDIGGNIR FGGTGINGGD VGYITGTINA QTMNPNSSHI 180
 TTGNSYADGG GTTLNFNATN NITINQASFD NSDAGTQKSY MNFKGSNIKI SGSSFTDDTN 240
 GGFNFSGNNN NSTISFNQTS FNQGTYNFSN SATLSFNNSN FNQGTYHFNS AQSTFENSNF 300
 NQGTYNFNDN TSFNNDTFNQ GTYNFNSSKV SFGSANTLNS SSPFASLKGS VSFNSGAIFN 360
 20 LNQTNNNNQ YDILTTNGAI QYGVYQSYLW LSGYQSGDKA ISHVEVSNNI YDVTFDINGQ 420
 DETLQETFSN QSIITQFLGD DLQQAQQTQ QEDVANSQNA LNKVASDNTI ANNDTSYTQS 480
 SNPTILKDAQ GLENTNQQTQ QDEKALEKDL AQIKQLANST TGFNEQAFTQ AQKQEQQDEQ 540
 ALQNDENAFN TEQEGLEQAI ANAKHANPTP NTPSPPTPTP IKHTAPNTPP SQVPPTPPSQ 600
 NLVYVRGTSN FQATLQKTYN NKGIYYIDPN LSGYQSGQSN TLSTYTANLL GRSFGVNANN 660
 25 GTLIIGNNTE SVNDNGLIWI GHGGFGYITG TFSAAANIYLT NNFKTGEGVS NSDGGGANIT 720
 FKASDNITMD GLNYNNAETV TKMIQTGASQ HSYTTFDATN NISVTDSDFS DMTWGKFSFS 780
 AKNISFSNAS FSGFTNPQGS STISTNASNS LSFTDSRLNG GAIYNLQANS LIPNNTQAVF 840
 NVLYSRGTSN FNATTQLLGN TSFTLSSQSL LNFNGDTTLQ NNANITLGNK SQAFAKNSLT 900
 LDNNSNLSLD NQSVLNANGT SAFNNQASLN IYNGSQAASF SLFFNGGTLN LNANSKLNAS 960
 30 SASFSNNTTI NLDDSVLNAN NTSSLNANIN FQASQADFG GNTTIDTASF NFDSASSLNF 1020
 NNLTANGALN FNGYAPSLTK ALMNVSGQFV LGNNGDINLS DINIFDNITK SVTYNINLNAQ 1080
 KGITLISGAN GYEKILFYGM KIQNATYSAN NNIQTWSFIN PLNSSQIIQE SIKNGDLTIE 1140
 VLNNPNSASN TIFNIAPELY NYQDSKQNPY GYSYDYSNQ AGTYYLTSNI KGLFTPKGSQ 1200
 TPQTPGTYSF FNQPLNSLNI YNKGFSSENL KTLGLILSQN SATLKEMIES NQLDNITNIN 1260
 35 EVLQLLDKIK ITQAQKQALL ETINHLTDNI NQTFNNGNLV IGATQDNVTN STSSIWFQGN 1320
 GYSSPCALDS ATCSSFRNTY LGQLLGSSTSP YLGYINADFK AKSIYITGTI GSSNAFESGG 1380
 SADVTFQSAN NLVLNKANIE AQATDNIFNL LGQEGIDKIF NQGNLANVLS QMAMEKIKQA 1440
 GGLGNFIENA LSPLSKELPA SLQDETLGQL IGQNNLDDL NNSGVMNEIQ NIISQKLSIF 1500
 GNFVTPSIIE NYLAKQSLKS MLDDKGLLNF IGGYIDASEL SSILGVILKD ITNPPTSLOK 1560
 40 DIGVVANDLL NEFLGQDVVK KLESQGLVSN IINNVISQGG LSGVYNQGLG SVLPPSLQNA 1620
 LKENDLGTLL SPRGLHDFWQ KGYFNFLSNG YVFNNSFSF NATGGSLNFV ANKSIIFNGD 1680
 NTIDFSKYQG ALIFASNGVS NINITTINAT NGLSLNAGLN NVSVQKEIC INLANCPTTK 1740
 NSSPANSSVT PTNESLSVHA NNFTFLGTII SNGAIDLSQV TNSVIGTLN LLENATLQAN 1800
 NLTTTNAFNN ASNSTANIDG NFTLNQQATL STNASGLNVM GNFNYSYGLV FNLSHSVSHA 1860
 45 IINTQGTATI MANNNPLIQF NASSKEVGTY TLIDSAKAIY YGYNNQITGG SSLDNYLKLY 1920
 ALIDINGKHM VMTDNGLTYN GQAVSVKDG LNVGFKDSQN QYIYTSILYN KVKIAVSNDP 1980
 INNPAQPTLK QYIAQIQGVQ SVDSIDQAG NQAINWLNKI FETKGSPLFA PYLLESHSTK 2040
 DLTTIAGDIA NTLEVIANPN FKNDATNILQ INTYTQMSR LAKLSDTSTF ARSDFLERLE 2100
 ALKNKRFADA IPNAMDVILK YSQNRVKNV VWATGVGGAS FISGGTGTLY GINVGYDRFI 2160
 50 KGVIVGGYAA YGYSGFHANI TQSGSSNVNV GVSRAFIK SELTMSLNET WGYNKTFINS 2220
 YDPLLSIINQ SYRYDTWTTD AKINYGYDFM FKDKSVIFKP QVGLSYYYIG LSGLRGIMDD 2280
 PIYNQFRANA DPNKKSFLT I NFALESRRHYF NKNSYFVIA DVGRDLFINS MGDKMVRFIG 2340
 NNTLSYRDGG RYNTFASIIIT GGEIRLFKTF YVNAGIGARF GLDYKDINIT GNIGMRYAF 2399

55 <212> Type : PRT
 <211> Length : 2399
 SequenceName : SEQ ID 92
 SequenceDescription :

60 Sequence

 <213> OrganismName : Helicobacter pylori J99
 <400> PreSequenceString :

65 MEIQQTTHRKI NRPLVSLVLA GALISAIPQE SHAAFFTTHVI IPAIVGGIAT GTAVGTVSGL 60
 LSWGLKQAEF ANKTPDKPDK VWRIQAGKGF NEFPNKEYDL YKSLSSKID GGWDWGNAAR 120
 HYWVKGQWNN KLEVDMKDAV GTYKLSGLRN FTGGDLVNM QKATLRLGQF NGNSFTSYKD 180
 SADRTTRVNF NAKNISIDNF VEINNRVSGS AGRKASSTVL TLQASEGITS SKNAEISLYD 240

	GATLNLASNS	VKLNGNVWVG	RLQYVGAYLA	PSYSTINTSK	VQGEVDFNHL	TVGDQNAQA	300
	GIIASNKTHI	GTLDLWQSAG	LNIIAPPEGG	YKDKPNSTTS	QSGTKNDKKE	ISQNNNSNTE	360
	VINPPNMTQK	TETEPTQVID	GPFAGGKDTV	VNIFHLNTKA	DGTIKVGGFK	ASLTNNAHL	420
	NIGKGGVNL	NQASGRILLV	ENLTGNITVD	GPLRVNNQVG	GYALAGSSAN	FEFKAGVDTK	480
5	NGTATFMNDI	SLGRFVNLKV	DAHTANFKGI	DTGNGGFNTL	DFSGVTDKVN	INKLITASTN	540
	VAVKNFNINE	LIVKTNGISV	GEYTHFSEDI	GSQSRINTVR	LETGTRSIFS	GGVKFKSGEK	600
	LVINDFYYS	WNYFDARNVK	NVEITRKFS	STPENPWGTS	KLMFNNTLGL	QNAVMDYSQF	660
	SNLTIQGDFI	NNQGTINYLV	RGGKVATLNV	GNAAMMFNN	DIDSATGFYK	PLIKINSAQD	720
	LIKNTAEHVL	KAKIIGYGNV	STGTNGISNV	NLEEQFKERL	ALYNNNNRMD	TCVVRNTDDI	780
10	KACGMAIGNQ	SMVNNPDNYK	YLIKAWRNI	GISKANGSK	ISVYVLGNST	PTENGNTTN	840
	LPTNTTNNAH	SANYALVKNA	PFAHSATPNL	VAIQHDFGT	IESVFELANR	SKDIDTLYTH	900
	SGAQGRDLLQ	TLIDSHDAG	YARQMIDNTS	TGEITKQLNA	ATDALNNVAS	LEHKQSGLOT	960
	LSLSNAMILN	SRLVNLRSKH	TNHINSFAQR	LQALKGQEFA	SLESAAEVLY	QFAPKYEKPT	1020
	NVWANAIGGA	SLNNGSNASL	YGTSAGVDAF	LNGNVEAIVG	GFSGYGYSSF	SNQANSLSNG	1080
15	ANNANFGVYS	RPFANQHEFD	FEAQGALGSD	QSSLNFKSTL	LQDLNQSYNY	LAYSATARAS	1140
	YGYDFAFFRN	ALVLKPSVGV	SYNHLGSTNF	KSNSQSQVAL	KNGASSQHLP	NANANVEARY	1200
	YYGDTSYFYL	HAGVLQEFAL	FGSNDVASLN	TFKINAARSP	LSTYARAMMG	GELQLAKEVF	1260
	LNLGVVYLHN	LISNASHFAS	NLGMRYSF				1288
	<212> Type : PRT						
20	<211> Length : 1288						
	SequenceName : SEQ ID 93						
	SequenceDescription :						
	Sequence						
25	-----						
	<213> OrganismName : Helicobacter pylori J99						
	<400> PreSequenceString :						
	MKKHILSLTL	GSLLVSTLSA	EDDGFYTSVG	YQIGEAQMV	TNTKGIQDLS	DRYESLNNLL	60
	NRYSLTNTLI	KLSADPSAIN	AVRENLGASA	KNLIGDKANS	PAYQAVLLAI	NAAVGFWNVV	120
30	GYVTQCGGNA	NGQKSISST	IFNNEPGYRS	TSITCSLNGH	SPGYGPMISI	ENFKKLNEAY	180
	QILQTALKRG	LPALKENNGK	VNVTYTYTCS	GDGNMNCSSQ	VTGVNNQKDG	TKTKIQTIDG	240
	KSVTTTSSK	VVDSRAGNT	TGVSYTEITN	KLEGVPDSAQ	ALLAQASTLI	NTINNACPYF	300
	HASNSSEANA	PKFSTTTGKI	CGAFSEIISA	IQKMITDAQE	LVNQTSVINE	HEQTPVGN	360
	NGKPFNPFTD	ASFAQGMLAN	ASAQAKMLNL	AEQVGQAINP	ERLSGTFQNF	VKGFLATCNN	420
35	PSTAGTGGTQ	GSAPGTVTQ	TFASGCAYVG	QTTNLKNSI	AHFGTQEQI	QQAENIADTL	480
	VNFKSSESEL	GNTYNSITTA	LSNIPNAQSL	QNAVSKKNP	YSPQIDTNY	YLNQNSYNQI	540
	QTTNQELGRN	PPRKVGIVSS	QTNGAMNGI	GIQVGKQFF	GQKRKGARY	YGFDPYNHAF	600
	IKSSFFNSAS	DVWTYGFAD	ALYNFINDKA	TNPLGKNNKL	SVGLFGGIAL	AGTSWLNSEY	660
	VNLATMNNVY	NAKMNVANFQ	FLFNMGVRMN	LARPKKKDS	HAAQHIGELG	LKIPTINTNY	720
40	YSFMGAELKY	RRLYSVYLN	VFAY				744
	<212> Type : PRT						
	<211> Length : 744						
	SequenceName : SEQ ID 94						
	SequenceDescription :						
45	Sequence						

	<213> OrganismName : Helicobacter pylori J99						
	<400> PreSequenceString :						
50	MIKKAKKFIP	FPLIGSLLAE	DNGWYMSVGY	QIGGTQQFIN	NKQLLENQNI	INSITQSAIN	60
	IAGPTTGLIT	LSSQTVIDAL	GYGVSNVTGN	QLEGISNIN	QIGKRKDFYS	SRQISSISQ	120
	IIGLKSSSDP	LKAHSSQITA	KLLSNTQSAF	DQGIALSSNI	ISAVNSLNPS	NNSQEVKAQL	180
	QNTAQSMAL	LQQIEHSITK	TTSTTYAQSL	LSNLTDVANA	SSNNTTYVSA	LVNALNTLGV	240
	GVFPTTSTH	VVLNPPGQV	FYPTNSLLGS	TSSNSNNQQQ	YNNLTLMNTL	QGELSTNNQN	300
55	NPNGCANQIQ	CLEQFIQNL	PLAATPTSTN	QANQQVQAI	QKLQSVAINA	LDNNAINNTT	360
	YNLNNLHNA	NFQAYQSTIE	QYNNALKQIS	WISFSEPKNL	LKNTSNNYQI	GTVTNDQGN	420
	ISAYDCTSAT	GSLSSDASSG	ISCSATSTTN	NTNSFSDNSLV	ATSKVQTING	KEQIGVNSFN	480
	LVSQVWSVYN	SLKTSEENLQ	KNAKILCNNG	SQSGTSPCNS	SSGGLSISGN	AQLQNILSPT	540
	NGTTTNTQAK	SNASKLKAMV	MVNNEEEAKT	TNPNQSSGPT	TQSSNSTVMG	ALNTVLQNVS	600
60	NFQQSIQSAF	QNOENNIQAW	ANALYNTSNP	NGNQSQNLTT	NNNQDLRIQL	RANFYQLINT	660
	INQQVPTDMN	ALINQSQQTQ	QTSGSASTTN	NACASGMGSS	GNWCYQQWSD	SKAYYSLQS	720
	ALGYQTQATT	QNGSSGGSNI	TYNVQQITLT	SGGLLNQIIT	NLKSVMGGSN	GGSSGNGTSQ	780
	INTAYQMLTD	ASDGKLGTYN	SSNSNSSNS	GNNNGYTPCN	STNGSNGTSG	SNCYEPNKQQ	840
	NATTATTTTD	SNLQKVYNDA	QKIANIASS	GNNKGVENGL	KQFFEALKSN	SSSLSNLCGN	900
65	GSSGSSSTCS	GGLINLLGAI	PTNGVSDTNN	LINLLTEFIK	TAGFIQNKDS	NVSTSLTSAF	960
	QAITSAISQ	FQALQNDISP	NAITLLQEI	TSNTTTIQSF	SQTLRQLLGD	KTFFMVQQL	1020
	IDAMINARNQ	VQNAQNQANN	YGSQPVLQY	AAAKSTQHGM	SNGLGVGIGY	KYFFGKARKL	1080

```
GLRHYFFFDY GFSEIGLANQ SVKANIFAYG VGTDFLWNLF RRTYNTKALN FGLFAGVQLG 1140
GATWLSSLRQ QIIDNWGNAN DIHSTNFQVA LNFGVRTNFA EFKRFAKKFH NQGVISQKSV 1200
EFGIKVPLIN QAYLNSAGAD VSYRRLYTFY INYIMGF 1237
<212> Type : PRT
5 <211> Length : 1237
    SequenceName : SEQ ID 95
    SequenceDescription :

Sequence
10 -----
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKQNLKPFKM IKENLMTQSQ KVRFLAPLSL ALSLSFNPVG ABEDGGFMTF GYELGQVVQQ 60
VKNPQGIKAE ELAGLLNSTT TNNTNINIAG TGGNVAGTLG NLFMMQLGNL IDLYPTLKTN 120
15 NLHQCGSTNS GNGATAAAAT NNSPCFQGNL ALYNEMVDSI KTLNQNISKI IFQGDNNTTS 180
ANLSNQSLSEL NTASVYLTYM NSFLNANNQA GGIFQNNNTQ AYENGVTAAQ IAYVLKQASI 240
TMGPSGDSGA AGAFLDAALA QHVFNANAG NDLSAKEFTS LVQNIIVNSQ NALTLANNAN 300
ISNSTGYQVS YGGNIDQARS TQLLNNTTNT LAKVTALNNE LKANPWLGNF AAGNSSQVNA 360
FNGFITKIGY KQFFGENKNV GLRYYGFFSY NGAGVGNNGPT YNQVNLLTYG VGTDLVLYNVF 420
20 SRSFGRSLN AGFFGGIQLA GDTYISTLRN SPQLASRPTA TKFQFLFDVG LRMNFGILKK 480
DLKSHNQHSI EIGVQIPTIY NTYYKAGGAE VKYFRPYSVY WVGYYAF 527
<212> Type : PRT
<211> Length : 527
    SequenceName : SEQ ID 96
    SequenceDescription :

Sequence
-----
<213> OrganismName : Helicobacter pylori J99
30 <400> PreSequenceString :
MKKTLTLLSL LSFGLHAEDD GFYASAGIRI GEAAQMVKNT KGIQQLSENY EKLNNLLNNY 60
NTLNTLVKLS SDPSAVNDAR DNLGSSTRNL LDVKANSPAY QAVLLALNAA VGLWQVTSYA 120
FTACGPGSNE NANGGIQTFN NVPQGNTTTI TCNSYYEPGH GGPISTKNYA IINKAYQIIQ 180
KALTANGEGI PVLNNTTTKL DFTINGDKRT GGEPNKKLVY PWSHGKAIST SWNATITAPT 240
35 TENINTNSA QELLKQASII ITTLNSACPN FQNGGSGYWA GISGNGTMCQ MFKNEISAIQ 300
GMIANAQEAQ AQAKIVSENT QNQNSLDAGK PFNPYTDASF AESMLKNAQA QAEILNQAEQ 360
VVKNFKEIPT AFVNDSLGVC YEYQGGERRG TNPGQTTSNT WGAGCAYVGQ TITNLKNSIA 420
HFGTQEQQIQ QAENIADTLV NFKSRYSELG NTYNSITTAL SNIPNAQSLQ NAVSKKNNPY 480
SPQGIDTNYI LNQNSYNQIQ TINQELGRNP FRKVGIVSSQ TNNGAMNGIG IQVGKQFFFG 540
40 QKRKGWARYY GFFDYNHAFI KSSFFNSASD VWTYGFADA LYNFINDKAT NFLGKNNKLS 600
VGLFGGIALA GTSWLNSEYV NLATMNNVYN AKMNVANFQF LFNMGVRMNL ARPKKKDSH 660
AAQHGIELGL KIPTINTNYI SFMGAELKYR RLYSVYLYNYV FAY 703
<212> Type : PRT
<211> Length : 703
    SequenceName : SEQ ID 97
    SequenceDescription :

Sequence
-----
50 <213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MIKKNRTLFL SLALCASISY AEDDGGFFTV GYQLGQVMQD VQNPQGAQSD ELARELNADV 60
TNNILNNNTG GNVAGALSNA FSQYLYSLLG AYPYTKLNGN VSANALLSGA VSGTCAAAG 120
TAGGTTLNTQ SACTAAGYYW LPSLTDRILS TIGSQTYNGT NTNFPNMQQQ LTYLNAGNVF 180
55 FNAMNKALEK NGTATANSTS STSGATGSDG QTYSQQAQIY LQGGQONILNN AANLLKQDEL 240
LLEAFNSAVA ANIGNKEFNS AAFTGLVQGI IDQSQLVYNE LTKNTISGSA VNNAGINSNQ 300
ANAVQGRASQ LPNALYNVQV TLDKINALNN QVRSMPLYLPQ FRAGNSRATN ILNGFYTKVG 360
YKQFFGKRN IGLRYYGFFS YNGASVGFRS TQNNVGLYTY GVGTDVLYNI FRSYQNRSV 420
DMGFFSGIQL AGETFQSTLR DDPNVKLHGK INNTHFQFLF DFGMRMNFQK LDGKSNRHNQ 480
60 HTVEFGVVVP TIYNTYKSA GTTVKYFRPY SVYWSYGYSF 520
<212> Type : PRT
<211> Length : 520
    SequenceName : SEQ ID 98
    SequenceDescription :

Sequence
-----
65
```

<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKKKFLSLTL GSLLVSAISA EDNGFFVSAG YQIGESAQMV KNTKGIQDLS DSYERLNNLL 60
TNYSVLNALI RQADPNAIN NARGNLNLSA KNLINDKNS PAYQAVLLAL NAAAGLWQVM 120
5 SYAISPCCPG KDTSKNGGVQ TFHNTPSNQW GGTITTCGTT GYEPGPYSIL STENYAKINK 180
AYQIIQKAFG SSGKDIPALS DTNTELKFTI NKNNGNTNTN NNGEEIVTKN NAQVLEQAS 240
TIITTLNSAC PWINNGGAGG ASSGSLWEGI YLKGDSACG IFKNEISAIQ DMIKNAIAV 300
EQSKIYAANA QNQRNLDTGK TFNPYKDANF AQSMFANAKA QAEILNRAQA VVKDFERIPA 360
EFVKDSLGVG HEVQNGHLRG TPSGTVTDNT WGAGCAYVGE TVTNLKDSIA HFGDQAERIH 420
10 NARNLAYTLA NFSSQYQKLG EHYDSITAAI SSLPDAQSLQ NVVSKKTNPV SPQGIQDNY 480
IDSNHSQVQ SRSQELGSPN FRRAGLIAAS TTNNGAMNGI GFQVGYKQFF GKNKRWGARY 540
YGFVDYNHTY NKSQFFNASS DVWTYGVGSD LLVNFINDKA TKHNKISFGA FGGIALAGTS 600
WLSQVYNLA NVNNYYKAKI NTANFQFLFN LGLRMLNARK KHRATDNAAQ HGIELGKIP 660
TINTNYSSL-GTTLQYRRLY SVYLNIVFAY 690
15 <212> Type : PRT
<211> Length : 690
SequenceName : SEQ ID 99
SequenceDescription :
20 Sequence

<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKIKKSLFAL SFLMASLSR AEDDGFYMSV GYQIGEAQVK VKNTGALQNL ADRYDNLNL 60
25 LNQYNLNSL VNLASTPSAI TGAIDNLSSS AINLTSATT SPAYQAVALA LNAAVGMWQV 120
IAFGISCPG PNLGPEHLEN GGVRSFDNTP NYSYNTGSGT TTTTCNGASN VGPNGILSSS 180
EYQVLNTAYQ TIQTALNQNG GGGMPALNSS KNMVNVINQT FTKNPTTEYT YPDGNGNYYS 240
GGSSIPIQLK ISSVNDANL LQQAATIIINV LTTQNPVNG GGGAWGFGGK TGNVMDIFGD 300
SFNAINEMIK NAQAVLEKTQ QLNANENTQI TQPDNPNPYT SKDTQFAQEM LNRANAQAEI 360
30 LSLAQVADN FHSIQGPIQQ DLEECTAGSA GVINDNTYGS GCAFVKETLN SLEQHTAYYG 420
NQVNQDRALS QTLNFKAL STLGNDSKAI NSGISNLPNA KSLQNMTHAT QNPNSPEGLL 480
TYSLDTSKYN QLQTVQELG KNPFRIGVI NYQNNNGAMN GIGVQAGYKQ FFGKKRNWGL 540
RYYGFFDYNH AYIKSNFFNS ASDVWTYGVG MDALYNFIND KNTNFLGKNN KLSVGLFGGF 600
ALAGTSWLNS QQVNLTMNG IYNANVSASN FQFLFDLGLR MNLARPKKKD SDHAAQHGM 660
35 LGVKIPTINT DYYSFMGAEL KYRRLYSVYL NYVFAY 696
<212> Type : PRT
<211> Length : 696
SequenceName : SEQ ID 100
SequenceDescription :
40 Sequence

<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
45 MKIKKSLFAL SFLMASLSR AEDDGFYMSV GYQIGEAQVK VKNTGALQNL ADRYDNLNL 60
LNQYNLNSL VNLASTPSAI TGAIDNLSSS AINLTSATT SPAYQAVALA LNAAVGMWQV 120
IAFGISCPG PNLGPEHLEN GGVRSFDNTP NYSYNTGSGT TTTTCNGASN VGPNGILSSS 180
EYQVLNTAYQ TIQTALNQNG GGGMPALNSS KNMVNVINQT FTKNPTTEYT YPDGNGNYYS 240
GGSSIPIQLK ISSVNDANL LQQAATIIINV LTTQNPVNG GGGAWGFGGK TGNVMDIFGD 300
50 SFNAINEMIK NAQAVLEKTQ QLNANENTQI TQPDNPNPYT SKDTQFAQEM LNRANAQAEI 360
LSLAQVADN FHSIQGPIQQ DLEECTAGSA GVINDNTYGS GCAFVKETLN SLEQHTAYYG 420
NQVNQDRALS QTLNFKAL STLGNDSKAI NSGISNLPNA KSLQNMTHAT QNPNSPEGLL 480
TYSLDTSKYN QLQTVQELG KNPFRIGVI NYQNNNGAMN GIGVQAGYKQ FFGKKRNWGL 540
RYYGFFDYNH AYIKSNFFNS ASDVWTYGVG MDALYNFIND KNTNFLGKNN KLSVGLFGGF 600
55 ALAGTSWLNS QQVNLTMNG IYNANVSASN FQFLFDLGLR MNLARPKKKD SDHAAQHGM 660
LGVKIPTINT DYYSFMGAEL KYRRLYSVYL NYVFAY 696
<212> Type : PRT
<211> Length : 696
SequenceName : SEQ ID 101
SequenceDescription :
60 Sequence

<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
65 MHKKVLLALT ASLICQESLF AKDKDYTLGK VSTAGKKDRS DYSGQVNLGY SGITAPKSWQ 60
DEEVKKYTG RTVISNKALT QQANQSIEEA LQNVPLQIR NATGVGAMPT IQIRFGAGG 120

SGHSDATLML VNGIPVYMAP YAHIELDIFP VTFQAIIDRID VIKGGGSVQY GPNTYGGIVN 180
IITKPIPNQW ENQAAERITY WAKARNAGFA APPDKTGDPS FIKSLGNNLL YNTYVRSGGM 240
INKHVGIIAQ ANWVRGQGFR DNSPSSISNY WLDGVYDINE SNGIKAYYQY YDFAIAQPGS 300
LSEQDYKINR FANLRPLNOK GGRSQRFGAV YENRFGDLDR VGGTFSFTYY GQLMTRDFQV 360
5 SSSYNSANMV TCFSEAACRA AGLPAGYNLA VPYYATNYNG WAEVENPVRS INNAFEPKVN 420
LIVNTGKVRQ TFIMGLRFMT TTFLQRQYLN TNECATKTS GAGFLCEGP NVMSGWKPHI 480
KHGVYRNWNN WRNNYTAVYL SDRIEAWDGR FFIVPGLRYA FVQYNNENAS NWMQIPEKDL 540
RKIKHMNNWM PSTNIGFIPV QGDHNVLTYP NYQSFVFPQ LDVLSYGGAE YFTQHFDTVE 600
AGARYTYKDK FSVNADYFRI WARDFATGQY SVYTSGPMKG NVRPINGYSQ GVELELYYRP 660
10 IRGLQFHAAF NYIDTRVTSH GPLTDLNGDV LKGTSYNKH FVSPFQFIF DARYNWRKTT 720
IGISSYFYSR AYSGISNSAA GGYGQMYY SGGNNYESVLN SGYQCEAWCM TQHEGLLPWY 780
WVWNIQVSQI FWENGRHRVT GSLQINNIFN MKYYFTGIGS SPAGLQPAPG RSVTAYLNYT 840
F 841
<212> Type : PRT
15 <211> Length : 841
SequenceName : SEQ ID 102
SequenceDescription :

Sequence
20 -----
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKKTLTLLSLS ASSLLNAEDN GFFISAGYQI GEAAQMVKN T GELKKLSDTY ENLSNLLTNF 60
NNLNQAVTNA SSPSEINAAI DNLKANTQGL IGEKTNPAY QAVYLALNAA VGLWNVIAYN 120
25 VQCGPGNSGQ QSVTFEGQPG HNSSSINCNL TGYNNGVSGP LSIENFKKLN QAYQTIQQAL 180
KQDSGFPVLD SAGKQVTITI TTQTNGANKS ETTTTTTTTN DAQTLLEQAS KMISVLTINC 240
PWVNHNNQGN GGAPWGLDTA GNVQCQVFATE FSAVTSMIKN AQEIVTQAQS LNQQNNQNAP 300
QDFNPYTSAD RAFAQNMLNH AQAQAKILEL ADQMKKDLNT IPSQFITNYL AACHNGGGTL 360
PDAGVTNNTW GAGCAYVBET ITALNNSLAH FGTQAEQIKQ SELLARTILD FRGSLSNLNN 420
30 TYNSITTTAS NTPNSPFLKN LISQSTNPNM PGGLQAVYQV NQSAYSQLLS ATQELGHNPF 480
RRVGLISSQT NNGAMNGIGV QVGYKQFFGE KRRWGLRYYG FFDYNHAYIK SSFFNSASDV 540
FTYGVGTDVL YNFINDKTTK NSKISFGVFG GIALAGTSWL NSQYVNLATF NNFYSAKMNV 600
ANFQFLFNLG LRMNLAKNKK KASDHAAQH GVELGVKIPTI NTNYYSLLTG QLQYRRLYSV 660
YLNYYVFAY 668
35 <212> Type : PRT
<211> Length : 668
SequenceName : SEQ ID 103
SequenceDescription :

Sequence
40 -----
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MRKLFIPLLL FSALEANENK GFFIEAGFET GLLEGTTQTQE KRHTTTKNTY ATYNYLPTDT 60
45 ILKRAANLFT NAEATSKLKF SSLSPVRVLY MYNGQLTIEN FLPPYLNNNVK LSFTDAQGNT 120
IDLGVIETIP KHSKIVLPGE AFDLKEAFD KIDPYTLFLP KFEATSTSTIS DTNTQRFVET 180
LNNIKTNLIM KYSNENPNNF NTCPPYNNNGN TKNDWCQNFT PQTAEEFTNL MLNMIAVLDS 240
QSWGDAILNA PFEFTNSSTD CSDSPSKCVN PGVNGRVDTK VDQYILNKQ GIINNFRKKI 300
EIDAVVLKNS GVVGLANGYG NDGEYGTGLV EAYALDPKKL FGNDLKTINL EDLRTILHEF 360
50 SHTKGYGHNG NMTYQRVPVT KDGQVEKDSN GKPKDSGLP YNVCSLYGGS NQPAFPSNYP 420
NSIYHNCADV PAGFLGVTA VVQQLINQNA LPINYANLGS QTYNLYNLASL NTQDLANSML 480
STIQKTFVTS SVTNHHFSNA SQSFRSPILG VNAKIGYQNY FNDFIGLAYY GIIKYNKYA 540
VNQKVQQLSY GGGIDLLLD IITYSNKNSP STGIQTKRNF SSGIFGGLR GLYNSYYVLN 600
KVKGSGNLDV ATGLNRYRYK SKYSVGISIP LIQRKASVVS SGGDYTNSFV FNEGASHFKV 660
55 FFNYGWVF 668
<212> Type : PRT
<211> Length : 668
SequenceName : SEQ ID 104
SequenceDescription :

Sequence
60 -----
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
65 MNKTTIKILM GMAILSSLQA AEAELDEKSK KPKFADRNTF YLGVGYQLSA INTSFSTSSI 60
DKSYFMTGNG FGVVLGGKFV AKTQAVEHV FRYGLFYDQT FSSHKSIST YGLEFSGLWD 120
AFNSPKMFLG LEFGLGIAGA TYMPGGAMHG IIAQYLKEN SLFQLLVKVG FRFGFFHNEI 180

TFGLKFPVIP NKKTEIVDGL SATTLWQRLP VAYFNYIYNF 220
<212> Type : PRT
<211> Length : 220
SequenceName : SEQ ID 105
5 SequenceDescription :

Sequence

<213> OrganismName : Helicobacter pylori J99
10 <400> PreSequenceString :
MKKTCKKTILL SLTLAASLLH AEDNGVFLSV GYQIGEAVQK VKNADKVQKL SDVYEQLSKL 60
LANDNGTSSK TSAQAINQAV NNLNESAKTL AGGTTNSPAY QATLLALRSA LGLWNSMGYA 120
VVCQGYIKKP GENNQKNFHY TDENGNGTTI NCGGSTNSNG THSPNGTNTL KADKNVSLSI 180
EQYEKIHEAY QILSKALKQA GLAPLNSKGE KLEAHVTTSK DQOQTSSDQT TTTTSVIDTT 240
15 NDAQNLLTQA QTIVNTLKDY CPMLIAKSSS NGGTNGANTP SWQTAGGGKN SCATFGAEFS 300
AISDMISNAQ KIVQETQQLN ANQPKNITQP NNFNLNSPGS LTALAQSMLEK NAQSQTEILK 360
LANQVADDFD KLSSGYLKDY IGKCDVSGVS SSNMTPQNMN TTWGGKGCAGV EETLTSLEKAS 420
TTDFNNQTPP QLDQAQTLAN TLTQELGNPP FKRVGIIIGSQ TNGGAMNGLG VQAGYKQFFG 480
QKRWRGLRYY GFFDYNHTYI KSSFFNSSSD VLTGTVGSDL LFNFINDKNT NFLGKNNKIS 540
20 VGLFPGGIALA GTSWLSNQFV NLKTTISNVYS AKVNTANFQF LFNGLRLTNL ARPKKKDSH 600
SAQHGMELGV KIPTINTNYY SYLGTKLEYR RLYSVYLYNVF FAY 643
<212> Type : PRT
<211> Length : 643
SequenceName : SEQ ID 106
25 SequenceDescription :

Sequence

<213> OrganismName : Helicobacter pylori J99
30 <400> PreSequenceString :
MKKTILLSLM VSSLFAENDG VYMSVGQIG EAAQMVKN TG EIQKVSNA YE NLNLLTRYN 60
ELKQTASNTD SSTAQAIDNL EKSASRLKTT PMTANQAVSS ALSSAVGMWQ VIASNLANS 120
LSSSEYEKLEK ATSQLLQNTL ENKNNNLKIE NDYDQLLTQA STIINTLQSQ CPGVDGGNGK 180
PWGINTSGNA CAIFGSTFNA INSMIDSAKK AAADARRTAP ESPNQNAFT NADFNKNLNQ 240
35 VSSVINDTIS YLKGDNLETI YNTIQKTPNS KGFQSLVSR S SYSYSLNETQ YSQFQTTTKE 300
FGHNPPFRSVG LINSQSNNGA MNGVGVQLGY KQFFGKKNKFF GIRYYGFFDY NYAYIKSNFF 360
NSASNVFTYG AGSDLLNFI NGGSDRNRKV SFGIFGGIAL AGTTWLNNQS ANLKITNSAY 420
SAKINNTNFQ FLFNTGLRLQ GIHHGIELGV KIPTINTNYY SFMGAKLAYR RLYSLYLYNV 480
LAY 483
40 <212> Type : PRT
<211> Length : 483
SequenceName : SEQ ID 107
SequenceDescription :

45 Sequence

<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MPKASQVLFF GAFLSTSLQG FEAKLNGFVD QSSTIGFNQH KINKERGIYP MQQFATIAGY 60
50 LGLGFSLLPK KVS DHVLK GK IGGMVGSIFY DGTKKFEDGS VAYNLFGYYD GFMGVYTNIL 120
QTDSLETQNM KHNKNVRNYV FSDAYLEYAY KNYFEIKAGR YLSTMPYKSG QTQGFQVSGQ 180
YKHARLTWFS SWGRAFYGS FLMDWFAART TYSGGFTKNN NGGYDSHGRK VLYGTHAVQL 240
TYKPHRFLIE GFYYLSPQIF NAPGVKIGWD SNPNFSGTGF RSDTAIGFF PIYYPWMIVK 300
SNGSPVYRYD TPATQNGQNL IIRQRFDINN YNVSIAFYKV FQANGWIGN MGNPSGVIMG 360
55 SNSVYAGFTG TALKRDAATI FLSCGTHFA KKFTWK FATQ YSNSVVSWEA RAMISLGYKF 420
TEYLSGSVDL AYYGVHTNKG FKPGENGPVP KNFPALYSDR SALYTALVAS F 471
<212> Type : PRT
<211> Length : 471
60 SequenceName : SEQ ID 108
SequenceDescription :

Sequence

65 <213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MLRLVSKTIC LSLISLFNPL EAFQKHQKDV FFVEAGFETG LLEGAQTKEQ AIAQNTQNTQ 60

KIYENPLTHP QTKEQPKEQN KSDTATPQSV YGRYYILQNT ILEKATELFT AANINGNGLT 120
FYSQNPVYVM AYNKDNAEFE GYGNNNSVVVI QNFLPYNLNN IELSYTDAQG KAVNLGVIET 180
IPKDSQIILP ASLFNNFSND SPFNSDGLQQ LQTTTTPFSD ANTQSLFEKL SQITTNLQMT 240
YENTDPFSSG NNDPNGPLAS PKPHYECPGY KKSCQVASVS FTPQTAEELT NLMLDMIAVF 300
5 DSKSWEEAVL NAPFQFSNSP SECGIDYPKC VNPFFNGLVD PKDEKYALTPEEVINSYRVA 360
NELTVNLLNA AKGFLGLGSQ LGSANAPDDD GFNQGVLGIA PFALDPEKLF GKNLNKVAIL 420
ALRDIIEHYG HTLGYTHNGN MTYQVRVLCQ EGNNGPEARCE GCHEVEKNGK EELEFSNGHE 480
VRDHDGYTYD VCSRFGGKNQ PAFPSNY PNS IYTNCAQVPA GLIGVTTAVW QQLINQNALP 540
INFANLNSQT SHLNAGLNAQ NFATSMVSAI AQNFSTSTT TYRSSSKNFR SPILGVNVKI 600
10 GYQHYFNDYT GLAYYGIQY NYAQANDEKI QQLSYGGMD VLFDFITTYT NKKQDHPTKK 660
VFASSFGVFG GLRGLYNSY VFNQVKSGN LDIVTGFNYR YKHSKYSIGV SVPLIQSGIK 720
IASNNGIYAD SVVLNEGGSF FKVFFNYGWV F 751
<212> Type : PRT
<211> Length : 751
15 SequenceName : SEQ ID 109
SequenceDescription :

Sequence

20 <213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MQNFVFNKKW LIYSSLLPLF FLNPLMAEDD GFFMGVSYQT SLAVQRVDNS GLNASQDAST 60
YIRQNAIALE SAAVPLAYYL EAMGQQTRVL MQMLCPDPSK RCLLYAGGVQ NGQNNNGDTG 120
NNPPRGVNA TFDMSQLVNN LNKLTQLIGE TLIRNPENLP NSKVFNKFG NQSTVIALPE 180
25 GLANTMDALN NDI TNALTTL WYNQTLTNKS FSTPSNTSVN FSPQVLQHLL QDGLATANN 240
QTICSTQMQC TATNEAKSIA QNAQNIFQAL MQAGILGGLA NEKQFGFTYN KAPNGSDSQ 300
GYQSFSGPGY YTKNDNTTQA PLKALPAGAT IGSGNGQYTY HPSSAVYYLA DSIIANGITA 360
SMIFSGMQNF ANKAAKLIGT SSYNQMQDAI NYGESLLSNT VAYGDFITNW VAPYLDLNNK 420
GLNFLPNYGG QLNAGANNQTP QLTPQQAQQE QKVIMNQLEQ ATNAPTPAQI NRILANPYSP 480
30 TAKTLMAYGL YRSKAVIGGV IDEMQTKVNO VYQMGFARNF LEHNSNSNNM NGFGVKMGYK 540
QFFGKKRMFG LRYYGFDYDFG YAQFGTESSL VKATLSSYGA GTDFLYNVFT RKRGTETADI 600
GFFAGIQLAG QTWKTNFLDQ VDG NHLKPKD TSFQFLFDLG IRTNFSKIAH QKRSRFSQGI 660
BFGLKIPVLY HTYYQSEGV T AKYRRDFSFY VGYNIGF 697
<212> Type : PRT
35 <211> Length : 697
SequenceName : SEQ ID 110
SequenceDescription :

Sequence

40 <213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKKTILLSLS LSLASSLLHA EDNGFFVSAG YQIGEAVQMV KNTGELKNLN EKYEQLSQYL 60
NQVASLKQSI QNANNIELVN SSLNYLKSFT NNNYNSTQS PIFNAVQAVI TSVLGFWSLY 120
45 AGNYLTFFVY NKDTQKPASV QGNPPFSTIV QNCSGIENCA MNQTTYDKMK KLAEDLQAAQ 180
QNATTKANNL CALSGCATTQ GQNPSSSTVSN ALNLAQQLMD LIANTKTAMM WKNIVIAGVS 240
NVSGAIDSTG YPTQYAVFNN IKAMIPILQQ AVTLSQSNHT LSASLQAQAT GSQTNPKFAK 300
DIYAFANQK QVISYAQDIF NLFSSIPKQD YRYLEKAYLK IPNAGKTPTN PYRQEVNLNQ 360
BIQTIQNNVS YYGNRVDAAL SVAKD VYNLK SNQTEIVTTY NNAKNLSQEI SKLPYNQVNT 420
50 KDIITLPYDQ NAPAAGQYNY QINPEQQSNL SQALAAMSNN PFKKVGMISS QNNNGALNGL 480
GVQVGYKQFF GESKRWGLRY YGFFDYNHGY IKSSFFNSSS DIWTYGGGSD LLVNFINDSI 540
TRKNNKLSVG LFGGIQLAGT TWLNSQYMN L TAFNNPYSAK VNASNQFLF NLGLRTNLAT 600
AKKKDSERSA QHGVELGIKI PTINTNYYSF LGTKLEYRRL YSVYLN YVFA Y 651
<212> Type : PRT
55 <211> Length : 651
SequenceName : SEQ ID 111
SequenceDescription :

Sequence

60 <213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MLKLASKTIC LSLISSFTAV EAFQKHQKDG FFIEAGFETG LLQGTQTQEQ TIATTQEKPK 60
65 PKPKPKPITP QSTYGYKYYIS QSTILKNATE LFAEDNITNL TFYSQNPVYV TAYNQESAEE 120
AGYGNNSLIM IQNFLPYNLN NIELSYTDDQ GNVVSLGVIE TIPKQSQIIL PASLFNDPQL 180
NADGFQQLQT NTTRFSDAST QNLFNKLSKV TTNLQMTYIN YNQFSSGNGS GSKPPCPPYE 240

NQANCVAKVP PFTSQDAKNL TNMLMLNMAV FDSKSWEDAV LNAPFQFSDN NLSAPCYSDY 300
LTCVNPYNDG LVDPKLIKLN KGDEYNIENG QTGSVILTPQ DVIYSYRVAN NIYVNLLPTR 360
GGDLGLGSQY GGPNGPGDDG TNFGALGILS PFLDPEILFG KELNKNVAIMQ LRDIIHEYGH 420
5 TLGYTHNGNM TYQVRMCEE NNGPBERCQG GRIEQVDGKE VQVFDNGHEV RDTDGSTYDV 480
CSRFKDKPYT AGSYPNISYT DCSQVPAGLI GVTSVAVWQQL IDQNALPVDF TNLSSQTNVL 540
NASLNTQDFA TTMLSAISQS LSSSKSSATT YRTSKTSRPF GAPLLGVNLK MGYQKYFNDY 600
LGLSSYGIK YNYAQANNEK IQQLSYGVGM DVLFDFITNY TNEKNPKSNL TKKVFTSSLG 660
VFGGLRGLYN SYLLNQQYKG SGNLNVGTGGL NYRYKHSKYS IGISVPLVQL KSRIVSSDGA 720
YTNSITLNEG GSHFKVFFNY GWIF 744

10 <212> Type : PRT
<211> Length : 744
SequenceName : SEQ ID 112
SequenceDescription :

15 Sequence

<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
20 MRKLFIPLLL FSALEANEKN GFFIEAGFET GLEGTQTQE KRHTTTKNTY ATYNYLPTDT 60
ILKRAANLFT NAEAISKLKF SLSLSPVRVLY MYNGQLTIEN FLPPYNLNNVK LSFTDAQGNV 120
IDLGVITIP KHSKIVLPGE AFDSLKIDPY TLFLEPKIEAT STSISDANTQ RVFETLNKIK 180
TNLVVNYRNE NKFKDHENHW EAFTPQTAE FTLNMLNMIA VLDSQSWGDA ILNAPFEFTN 240
SPTDCDNDPS KCVNPGTNGL VNSKVDQKYV LNKQDIVNKE KNKADLDVIV LKDSGVVGLG 300
SDITPSNNDG GKHYGQLGVV ASALDPKKLF GDNLKTINLE DLRTILHEFS HTKGYGHNGN 360
25 MTYQRVPVTK DGQVEKDSNG KPKDSGLPY NVCSLYGGSN QPAFPSNYPN SIYHNCADVP 420
AGFLGVTAAB WQQLINQNAL PINYANLGSQ TNYNLNASLN TQDLANSMLS TIQKTFVTSS 480
VTNHHSNAS QSFRSPILGV NAKIGYQNYF NDFIGLAYYG IIKYNYAKAV NQKVQQLSYG 540
GGIDLLLDLI TTYSNKNSTP GIQTKRNFSS SFGIFGGLRG LYNYSYVLNK VKGSGNLDVA 600
TGLNRYKHS KYSVGISIP LQRKASVVSS GGDYTNSTFV NEGASHFKVF FNYGWVF 657

30 <212> Type : PRT
<211> Length : 657
SequenceName : SEQ ID 113
SequenceDescription :

35 Sequence

<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
40 MSLATSYNVS NNFSKFNKR VRGYLICLVC NTPKMIQRL NGVSFYGCSD YVNGKDCKGV 60
LREINGSMKM VCLHCENTPI MEKVESGRGG AYACKNCNRK FYFIDLAKQN ERKKOLEKEK 120
KELLNKIEKQ KIKHLERFIL AGVKANIKEN SFPLGCKNYP KCEWTASMD S QDLKCPKCNR 180
LMKRKNFKN NEFFTATSLT LNAIEFCLYI NLKKKETNV 219

45 <212> Type : PRT
<211> Length : 219
SequenceName : SEQ ID 114
SequenceDescription :

50 Sequence

<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
55 MEIKKYFLYA LFFLLFSGLF LSKLQAYKFN MSIVGKVSSY TKFGFNNQRY QPSKDIYPTG 60
SYTSLGELN LSMGLYKGLR AEVGAMMAAL PYDSTAYQGN NIPNGQPGSR TDPFGAGIFW 120
QYIGWYAGHS GLNVQKPLA MVHNAFLSYN YKDKFSFGV KGGRYDAEY DWFTSYTQGV 180
EGFVKYKDR LRVMSDARA SASSDWFYF GRYTSGKAL MIADLKYEK NLKINPYFYA 240
IFQRMYPGI NITYDTNPNF NNGGFRFVGT FVGFFPIFAT PANQNDIILF QQVPLGKSGQ 300
TYFFRTRFY NKWQFGGSVY KNIGNANGDI GIYGDPLGYN IWTNSIYDAE INNIVGADVI 360
NGFLYVGSQY RGSWVKILGR WTDSPRADER SLALFLSYFS NKYNIRMDLK LEYYGNITKK 420
60 GYCIGYCGMY VPDVDPNGPGT QPLTHNVYS RSHIMFNITY GFRIY 465

<212> Type : PRT
<211> Length : 465
SequenceName : SEQ ID 115
SequenceDescription :

65 Sequence

<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKKTILLSLS LSLASSLLHA EDNGFFVSAG YQIGEAQVMV KNTGELKNLN DKYEQLSQSL 60
AQLASLKSKI QTANNIQAVN NALSDLKSFA SNNHTNKETS PIYNTAQAVI TSVLAFWSLY 120
5 AGNALSFHVT GLNDGGSNSPL GRIHRDGNCT GLQQCFMSKE TYDKMKTLAE NLQKAQGNLC 180
ALSECSSNQS NGGKTSMTTA LQTAQQLMDL IEQTKVSMVW KNIVIAGVTN KPNGAGAIT 240
TGHVTDYAVF NNIKAMPLIL QQALTLSQSN HTLSTQLQAR AMGSQTNREF AKDIYALQN 300
QKQILSNASS IFNLFNSIPK DQLKYLENAY LKVPHLGKTP TNPYRQNVNL NKEINAVQDN 360
VANYGNRLDS ALSVAKDVYN LKSNQTEIVT TYNDAKNLSE BISKLPYNQV NVTNIVMSPK 420
10 DSTAGQYQIN PEQQSNLNQA LAAMSNNPFK KVGMISSQNN NGALNGLGVQ VGYKQFFGES 480
KRWGLRYYGF FDYNHGYIKS SFFNSSSDIW TYGGGSDLLV NFINDSITRK NNKLSVGLFG 540
GIQLAGTTWL NSQYMNLTAF NNPYSAKVNA SNFQFLFNLG LRTNLATAKK KDSERSAQHG 600
VELGIKIPTI NTNYSYFLGT KLEYRRLYSV YLNYVFAY 638
<212> Type : PRT
15 <211> Length : 638
SequenceName : SEQ ID 116
SequenceDescription :

Sequence

20 <213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKLKKRKVAA TLLKRLTLPL LFTTGS LGAV TYEVHGDFIN FSKVGFNRSP INPVKGIYPT 60
ETFVNLTGKL EGSVHLGRGW TVNVGGVLGG QVYDNTRYDR WAKDFTPPSY WDKTSCGTDS 120
25 LSLCMNATKM WQQQPGGGII DPERGIGYMYM GEWNGLFPNY YPANAYLPGH SRRYEVYKAN 180
LTYDSDRVHM VMGRFDVTEQ EQMDWIYQLF QGFYGTFFKLT KNMKFLLFSS WGRGIADGQW 240
LFPIYREKPW GIHKAGIYR PTKNLMIHYP VYLIPMVGTL PGAKIEYDTN PEFSGRGIRN 300
KTTFFVLYDY RWNNAEYGRY APARYNTWDP FLDNGKWRGL QGPGGATLYL HHHIDINNYF 360
VVGAYLNLG NPNMNLGTWG NPVALDGIEQ WVGGIYSLGF AGIDNITDAD AFTEYVKGEG 420
30 KHGKFSWSVY QRFTTAPRAL EYGIGMYLDY QFSKHVKAGL KLVWLEFQIR AGYNPGTGFL 480
GPNQPLNLN NGLFESSAFA QGPQNMGGIA KSITQDRSHL MTHISYSF 528
<212> Type : PRT
<211> Length : 528
SequenceName : SEQ ID 117
SequenceDescription :

Sequence

40 <213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKNFSPLYCL KKLKKRHIA LSLPLLSYAN GFKIQEQSLN GTALGSAYVA GARGADASFY 60
NPANMGFTND WGENRSEFEM TTTVINIPAF SFKVPTTNQG LYSVTSLEID KSQQNILGII 120
NTIGLGNILK ALGNTAATNG LSQAINRVQG LMNLTNQKV V TLASKPDTQI VNGWTGTNPF 180
VLPKFFFYKTR THNGFTFGGS FTAPSGLGKMK WNGKGGEFLH DVFIMMVELA PSMSYTINKR 240
45 FSVGVGLRGL YATGSFNNTV YVPLEGASVL SAEQILNLPN NVFADQVPSN MMTLLGNIGY 300
QPALNCQKAG GDMSDQSCQE FYNGLKKIMG YSGLIKASAN LYGTQVQVQK SNGQGVSGGY 360
RVGSSLRVFD HGMFVSVYNS SVTFNMKGGL VAITELGPSL GSVLTKGSLN INVSLPQTLS 420
LAYAHQFFKD RLRVEGVFER TFWSQGNKFL VTPDFANATY KGLSGTVASL DSETLKKMVG 480
LANFKSVMNM GAGWRDTNTF RLGVTYMGKS LRLMGALDIDY QAPSPQDAIG IPDSNGYTVA 540
50 FGTKYNFRGF DLGVAGSFTF KSNRSSLYQS PTIGQLRIFS ASLGYRW 587
<212> Type : PRT
<211> Length : 587
SequenceName : SEQ ID 118
SequenceDescription :

Sequence

55 <213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
60 MAPQVNTNIN AMNAHVQSAL TQNALKTSLE RLSSGLRINK AADDASGMTV ADSLRSQASS 60
LQQAIAANTND GMGIIQVADK AMDEQLKILD TVKVQATQAA QDGQTESRK AIQSDIVRLI 120
QGLDNIGNTT TYNGQALLSG QFTNKEFQVG AYSNQSIKAS IGSTTSKIG QVRIATGALI 180
TASGDISLTF KQVDGVNDVT LESVKVSSSA GTGIGVLAEV INKNSNRTGV KAYASVITTS 240
DVAVQSGSLS NLTLNGIHLG NIADIKKND S DGRLLVAINA VTSETGVEAY TDQKGRNLNR 300
65 SIDRGIEIK TDSVSNGPSA LTMVNGGQDL TKGSTNYGRL SLTRLDAKSI NVVSASDSQH 360
LGFTAIGFGE SQVAETTVNL RDVTGNFNAN VKSASGANYN AVIASGNQSL GSGVTTLRGA 420
MVVIDIAESA MKMLDKVRSD LGSVQNQMIS TVNNISITQV NVKAAESQIR DVDFAEESAN 480

FNKNNILAQS GSYAMSQANT VQQNILRLLT 510
<212> Type : PRT
<211> Length : 510
SequenceName : SEQ ID 119
5 SequenceDescription :

Sequence

<213> OrganismName : Helicobacter pylori J99
10 <400> PreSequenceString :
MAGTQAIYES SSAGFLSQVS SIISSTSGVA GPFAGIVAGA MTAALPIVV GFTNPQMTAI 60
MTQYNQSIAS AVSVPKMAAN QQYNQLYQGF NDQSMVAVGNN IILNISKLTGE FNAQGNTQSA 120
QISAVNSQIA SILASNTTPK NPSAIEAYAT NQIAVPSVPT TVEMMSGILG NITSAAPKYA 180
LALQEQLRSQ ASNSSMNDTA DSLDSC TALG ALVGSSKVFF SCMQISMTPM SVSMPTVYAK 240
15 YQAVATKALT SGVNPMTTPA CPIGDKVLAV YCYAEKVAEI LREYYIEFVK NNTNLLQNAS 300
QMILNQSGLA TSTYDTQAIS NISSLYNYNI VANKSFLKSH LTYLDYIKDK LKGQKDSYLT 360
ERVQTKLIVK 370
<212> Type : PRT
<211> Length : 370
20 SequenceName : SEQ ID 120
SequenceDescription :

Sequence

25 <213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MTNEAINQQP QTEAAFNPQQ FINNLQVAFI KVDNVVASFD PNQKPIVDKN DRDNRQAFEK 60
ISQLREEFAN KAIKNPTKKN QYFSSFISKS NDLIDKDNLI DTGSSIKSFQ KFGTQRYQIF 120
MNWVSHQNDP SKINTQKIRG FMENIIQPPI SDDKEKAEFL RSAKQAFAGI IIGNQIRSDQ 180
30 KFMGVFDESL KERQEAENKNG EPNGDPTGGD WLDIFLSFVF NKKQSSDLKE TLNQEPVPHV 240
QPDVATTTTD IQSLPPEARL LLDERGNFSK FTLGDMNMLD VEGVADIDPN YKFNQLLIHN 300
NALSSVLMSG HNGIEPEKVS LLYGNNGGPE ARHDWNATVG YKNQRGDNVA TLINVHMKNK 360
SGLVIAGGEK GINNPSFYLY KEDQLTGSQR ALSQEEIQNK VDFMEFLAQN NAKLDNLSKK 420
EKEKFQNEIE DFQKDSKAYL DALGNDHIAF VSKKDKKHLA LVAEFGNGEL SYTLKDYGKK 480
35 ADKALDREAK TTLQGS LKHD GVMFVDYSNF KYTNASKSPD KGVGATNGVS HLEAGFSKVA 540
VFNLPLNLNL AITSVVRQDL EDKLIAGLS PQEANKLVKD FLSSNKELVG KALNFNKAVA 600
BAKNTGNYDE VKQAQKDLEK SLKKRERLEK DVAKNLESKS GNKNKMEAKS QANSQKDEIF 660
ALINKEANRD ARAIAYAQNL KGIKRELSDK LENINKDLKD FSKSFDEFKN GKNKDFSKAE 720
ETLKALKGSV KDLGINPEWI SKVENLNAAL NEFKNGKNKD FSKVTQAKSD LENS IKDVII 780
40 NQKITDKVDN LNAQVSVAKA TGDFSGVEQA LADLNFSKE QLAQQAQKNE DFNTGKNSAL 840
YQSVKNGVNG TLVGNGLSKA EATTLNFSK DIKKELNAKL GNFNNNNNNG LENSTEPIYT 900
QVAKKVKAKI DRLDQIASGL GDVGQAASFL LKRHDKVDDL SKVGLSANHE PIYATIDDLG 960
GPFPLKRHDK VDDL SKVGLS REQKLTQKID NLNQAVSEAK ASHFDNLDQM IDKLKDS TKK 1020
NVVNLVYVESA KKVPTSLSAK LDNYATNSHT RINSNVKNGT INEKATGMLT QKNSEWLKLV 1080
45 NDKIVAHNVG SAPLSAYDKI GFNQKNMKDY SDSFKFSTRL SNAVKDIKSG FVQFLTNI FS 1140
MGSYS LMKAS VEHGVKNTNT KGGFQKS 1167
<212> Type : PRT
<211> Length : 1167
SequenceName : SEQ ID 121
50 SequenceDescription :

Sequence

55 <213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKTNGHFKDF AWKKCFLGAS VVALLVGCSP HIIETNEVAL KLVNHPASEK VQALDEKILL 60
LRPAFYSDN IAKEYENKFK NQTTLKVEEI LQNQGYKVIN VDSSDKDDFS FAQKKEGYLA 120
VAMNGEIVLR PDPKRTIQKK SEPGLLFSTG LDKMEGVLP AGFVKVTILE PMSGESLDSF 180
TMDLSELDIQ EKFLKTTHSS HSGGLVSTMV KGTDNSNDAL KSALNKIFAS IMQEMDKKLT 240
60 QRNLESYQKD AKELKNKRN 260
<212> Type : PRT
<211> Length : 260
SequenceName : SEQ ID 122
SequenceDescription :

Sequence

65

<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MKSCLKLKYR LLFLPLLLPLG TSLANTYLL QDHNTLTPYT PFTTPLNGGL DVVRAAHLHP 60
SYELVDWKRV GDTKLVALVR SALVRVKFQD TTSSDQSNNT QNALSFDTQE SQKALNGSQS 120
5 GSSDTSQSNQ QDFASYVLIF KAAPRATWVF ERKIKLALPY VKQESQSGSD QGSNGKGSly 180
KTLQDLLEQ PVTPTPNAG LARVNGVAQD TVHFGSGQES SWNSQRSQKG LKNNPGPKAV 240
TGFKLDKGRA YRKLNESWPV YEPLDSTKEG KGDDESSWKN SEKTTAENDA PLVGMVGSAG 300
AGSASSLQGN GSNSSGLKSL LRSAPVSVPP SSTSNQTLSTL SNPAPVGPQA VVSQPAGGAT 360
AAVSVNRTAS DTATFSKYL N TAQALHQMVG IVPGLEKWGG NNGTGVVASR QDATSTNLPH 420
10 AAGASQTGLG TGSPREPALT ATSQRVAVTV AGPLRAGNSS ETDALPNVIT QLYHTSTAQL 480
AYLNGQIVVM GSDRVPSLWY WVVGEDQESG KATWWAKTEL NWGTDKQKQF VENQLGFKDD 540
SNSDSKNSNL GAQGLTQPAY LIAGLDVVD HLVAFAFKAG AVGYDMTTDS SASTYNQALA 600
WSTTNGDSD KKYKALVENT AGLNGPINGL FTLLDFTFAY TPVSGMKGGS QNNEEVQTTY 660
FVKSDQKATA KIASLINASP LNSYGGDGVV VFDALGLNFN FKLNEERLPS RTDQLLVYGI 720
15 VNESELKSAR ENAQSTSDDN SNTKVWNTNT ASHYLPVPY YSANFPEAGN RRRAEQRNGV 780
KISTLESQAT DGFANSLNLF GTGLKAGVDP APVARGHKPN YSAVLLVRGG VVRLNFPNDT 840
DKLLDSTDKN SEPISFSYTP FGSAESAVDL TTLKDVITYA ESGLWFYTFD NGEKPTYDGK 900
QQQVKNRKY AVITVSRGTI EFNEDANTTT LSQAPALAV QNGIASSQDD LTGILPLSDE 960
FSAVITKQT WTGKVDIYKN TNGLFKDDQ LSENVKRRDN GLVPIYNEGI VDIWGRVDF 1020
20 ANSVLQARNL TDKTVDEVIN NPDILQSFFK FTPAFDNQRA MLVGEKTSST TLTVPKPKIEY 1080
LDGNFYGEDS KTAGIPLNID FPSRIFAGFA ALPSWVIPVS VGSSVGILLI LLILGLGIGI 1140
PMYKVRKLQD SSFVDVFKKV DTLTTAVGSV YKKIITQTSV IKKAPSALKA ANNAAPKAPV 1200
KPAAPTAPRP PVQPPKKA 1218
<212> Type : PRT
25 <211> Length : 1218
SequenceName : SEQ ID 123
SequenceDescription :

Sequence
30 -----
<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MHQTKKTALS KSTWILILTA TASLATGLTV VGHFTSTTTT LKRQQFSYTR PDEVALRHTN 60
AINPRLTPWT YRNTSFSSLP LTGENPGAWA LVRDNSAKGI TAGSGSQQT YDPTREAAAL 120
35 TASTTFALRR YDLAAGRALYD LDFSCLNPQT PTRDQTGQIT FNPFGGFGLS GAAPQWNEV 180
KNKVPVEVAQ DPSNPYRFAV LLVPRSVVYY EQLRGLGLP QORTESGQNT STTGAMFGLK 240
VKNAEADTAK SNEKLQGAEA TGSSTTSQSG DTKVKALKIE VKKSDSEDN 300
GQLQLEKNDL ANAPIKRSEE SGQSVQLKAD DFGTALSSSG SGGNSNPGSP TPWRPWLATE 360
QIHKDLPKWS ASILILYDAP YARNRTAIDR VDHLDPKAMT ANYPPSWRTP KWNHHGLWDW 420
40 KARDVLLQTT GFFNPRRHE WFDGGQTVAD NEKTGFVDVN SENTKQGFQK EADSDKSAPI 480
ALPFEAYFAN IGLNTWFGQA LLVFGNGHV TKSHTAPLS IGVRFRVYNA TGTSATVTGW 540
PYALLFSGMV NKQTDGLKDL PFNNNRWFY VPRMAVAGAK FVGRVLVLAG TITMGDTATV 600
PRLLYDELES NMLNVAQGGQ LLREDLQLFT PYGWANRPDL PIGAWSSSSS SSHNAPYYFH 660
NMPDWQDRPI QNVVDAFIK WEDKNGKDDA KYIYPYRSG MWAQVYNWS NKLTQPLSA 720
45 DFNENAYQP NSLFAAILNP ELLAALPDKV KYGKENEFAA NEYERFNQKL TVAPTQGTNW 780
SHFSPILSRF STGFNLVGSV LDQVLDYVPW IGNGYRYGNN HRGVDDITAP QTSAGSSSGI 840
STNTSGRSF LPTFSNIGVG LKANVQATLG GSQTMITGGS PRRTLDQANL QLWTGAGWRN 900
DKASSGQSDE NHTKFTSATG MDQQGQSSTS AGNPDLSLQD NISKSGDSLIT TDGNAIDQQ 960
EATNYTNLPP NLPTADWP N ALSFTNKNN QRAQLFLRGL LGSIPVLVNR SGSDSNKFQA 1020
50 TDQKWSYTDL HSDQTKLNL AYGEVNGLLN PALVETVYFGN TRAGGSGSNT TSSPGIGFKI 1080
PEQNNDKAT LITPGLAWTP QDVGNLVVSG TTVSFQLGGW LVTFTDFVKP RAGYLGGLT 1140
GLDASDATQR ALIWAPRPWA AFRGSWVNR GRVESVWDLK GVWADQAQSD SQGSTTTATR 1200
NALPEHPNAL AFQVSVVEAS AYKPNSSGQ TQSTNSSPYL HLKPKKVTQ SDKLDDDLKN 1260
LLDPNQVVRTK LRQSFQTDHS TQFPQSLKT TTPVFGTSSG NLSSVLSGGG AGGGSSGSGQ 1320
55 SGVDLSPVEK VSGWLVGQLP STSDGNTSST NNLAPNTNTG NDVVGVGRLS ESNAAKMND 1380
VDGIVRTPLA ELLDGEQTA DTGPQSVKFK SPDQIDFNRL FTHPVTDLFD PVTMLVYDQY 1440
IPLFIDIPAS VNPKMVRKLV LSFDTNEQSL GLRLEFFKPD QDTQPNNNVQ VNPNNGDFLP 1500
LLTASSQGPQ TLFSPFNQWP DYVLPLAITV PIVVIVLSVT LGLAIGIPMH KNKQALKAGF 1560
ALSNQKVDVL TKAVGSVFKE IINRTGISQA PKRLKQTSAA KPGAPRPPVP PKPGAPKPPV 1620
60 QPPKKPA 1627
<212> Type : PRT
<211> Length : 1627
SequenceName : SEQ ID 124
SequenceDescription :

Sequence
65 -----

<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MGYKLRWPL VAFFTFTGIGL GVVLAACSAL NTSNLFPRQN RSKQLIGFTE NNIIKPEAVL 60
KALAEDNGT ETILRVNFGE ALKSWYQNNK DRNIATRLTI FSENVEDEHD NLLDQKQAAE 120
5 PINWPIELQK EYDQWGGSES SWKALKLYDR LIADFQSLIF SNIVANVQLT DGSDQFKPTT 180
KDNLDSTSNK IKFVNSKPND PNGEFFANLQ AYLFAQWVVE ENPLPLTQAF FAYQAPKDGL 240
DSLYDQAAIG SALQLGYAFP AFREPNNGQS QGKTTFDPTP NSAQNFGDFI KAVFPEQKNG 300
QTQSSNTSSR TGLFDWQTKW NTNGAANKLL VTKSNLRGAF KGVGLATATII DQYEYLVGGS 360
KTSSLPEVKV DSNKSNQNPL DSFFMEGKDA VAIRSIVSRA KIAMTDQTPG FKVNPAPVKV 420
10 KQSQQNDTFY QNQRKLSGGQ SGNNSQGHK HYLQDAVRLT SSQAMAAAST GADSSSGTNN 480
GGSSGGNSVL IPLPRSAALT HTQQQVQQT STLTQTPVYAR GDDGTALAI DGGDYFLANN 540
KRDFTKQADI LLYRYLQAKS NNFKENGVEF SLNLLSGLS FQTWAQTGLT AKLYGALVAM 600
MGSQGQTQVK GSVQGSRAA SVSQTTQQN RQSTDTQES EVVKLAKSL KSSADLAKPF 660
TDNPTFKKAL TDIQSEYKYD LAAAGKLSEF KKLGEVSGL QQAIIIDRADK YIQLEKQAQK 720
15 SAIGLQPLP YQRASDGSYP ALEKFFIPED SAADGKVKAS ESGSAALVTL KTTDSQKSTN 780
TVKQPDIKPT RENNDKKLKQ LTSDEVTKAS SLITKWGATP QIGSQFSEIV SLKSKDNKPQ 840
TMILALLSD VGIKWTILN SFKEWFFTNT NDFKNYDSE KKLKGNEYK DFNDLVKQTL 900
YLRWQRLTS KEKFGYKEL GSVKAQAAQS GMSVLSSSAA VANAVASSGM QKSGDQTLLE 960
LGKKAFFESL EASSSDGQYK YLRFLSTLMW LVKDGAKNYK RLLQQAITVG TRAFVSWTVS 1020
20 YDDTATASAA AAKAQVAVLK TAQATNTQSD NPFNKVFQNP DYVQGSSETNW FNDKSTPIKP 1080
DSLLESESTY NFTAEPFDDK TKSQKRSTGG TTNEKHFFGF NGLTINSPQS VSTASAGLTE 1140
QIFNNFGQLV TSSDKSGALS QYKDKATLKR LIQNTNSDAE LNAFGEVLHR AVNVDTSNLG 1200
RPNSSGEPLI SFDNKKKFLV DVVDKLDDVY FNKFEYVVG TKVKMSDSSS SSQGTKTIRK 1260
PKPHHSRTR VSRLWAMSR LPRTLTKFL LVEKLIRTVL 1300
25 <212> Type : PRT
<211> Length : 1300
SequenceName : SEQ ID 125
SequenceDescription :
30 Sequence

<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MKKLLIKPQF WFLTLGGFIS SSVILVACAT PSNSALQTVF KARSNQFFNG EQGSLQNALA 60
35 TALKDPEANK QFVAAPLLKA LTAWYENNQD KQVTQFFKDT KKSVDQYQY AVDKVVSASR 120
NKNLFVQQLD LDSAGGVRNL KSPEVVWTAH 150
<212> Type : PRT
<211> Length : 150
SequenceName : SEQ ID 126
SequenceDescription :
40 Sequence

<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MQQQGETKDQ YNTFGLRLVR NSVGVSVLGL DGFVKFIKGG SGGSNNGSSS AKKIDKEEQK 60
KFLKFRAPQA KIGTFYNTNF AFSFPLNETL KGWFDKHRGL ILANALVKVT LDTKEKASKA 120
LVDAFSSYKN WLSEYTPVGL ATTMISFYFD QMKALNNKLL ERVRSNLQNV NQANPTPWLN 180
GLSAKLPIYV TNGNYEKLNN YFTFLITKVL WPKVGTEDTN VSEKSKLKT KTEDVNKIRE 240
50 KILNNIDSKL KTFVQKLKPT LAPRPAYSNV ILLNINNDKV WSAGANWSLA VLLDPKKVNP 300
LSFMLLKQMF DQNSLFKKAK TLFENIQNKA KTSKSGSGST TTNDADALS KVIGNYYYNT 360
WAKLTDKSIY GNLKDDKFDD LFKLAFDSSI NEKSFNVDYK AVIEHYRFIY TLEWLVDKNL 420
KNFKDLLKAN LKFGIEAFIA YKNTETQNF NPKGIFGSYF NYENETNAAK SATQIIDPNS 480
FFYKTTTKPE AKTTQSANTA VMVQNTQMN QQTNSYGFTE LSTSSGSM LG AATQQAILDQ 540
55 ITKTSLQQYG SQADLKKIIG ETKNQLLLDR IANQLIALKP NTSGNSGTQK TIAAYFQTD 600
VGNPTLDFKA KQKLLLDVLD QYKDFFGNNA QAVQRDSGKS GTGNYLTYTD GSDKITYLQF 660
SYKDIDGLSL SSSNGTSSKF ASDVVAALLL FQAAYKGTQ LALSSINKPQ LPIGDKRIKT 720
GIDLLK 726
<212> Type : PRT
60 <211> Length : 726
SequenceName : SEQ ID 127
SequenceDescription :
Sequence

65 <213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :

5 MKSFLRKPKF WLLLLGGLST SSILSACAT PSNSALQAVF KPTSNQFFNG EHGTIQSALN 60
TALRDPETNK KFVAAPLLKA LEAWYENNQD KNTQFLKDT KTNVDNQYKT VVDKVVSAPR 120
NKSLFVQDDL LDSSGGSEAT WKARKLFEQL ISDFASRVFQ KNYLSYKENG KVSAGPFLYD 180
TISKNSNWQN IVFDAVNFPF TNDDFFAKIQ SEVFDQWAEY TDPTIISVT LKYSAPN 237

<212> Type : PRT
<211> Length : 237
SequenceName : SEQ ID 128
SequenceDescription :

10 Sequence

<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :

15 MINFLFNQMN ALNNKFLERA KALNQNVNQA NPTPWNLGLS AKLPYVRTNG NYEKLNNYFT 60
FLIVKYMWK VGNEDASLSK DSSINKLTK TEDVNKIRDK ILEDIQKKVQ EFVKNKLKPT 120
LAPRQTYSNV ILLNVNNDKV WSMGANWALA NLLDTSKINP LSFMLLKQTF DQNDLFKKAK 180
KLFEDIQSKT NGSSSGGMQG SNTSSSEGAD ALSKVIGNY YNSWAKLTDK SIYGNPKDNK 240
FDDLFLKAFE DSINEKSFNV DYKAVIEHYR FIYLEWLNV GNLKNFKDLL KANLKFGEIA 300
20 FIAYKNTETK EFSNPQGVFG SAFNYENETN EVKIAAQNLD PNNFFYKTTT KPEEVKTAQN 360
GASMMVMQKN MQSTMQDSNH YGFTGLNST SSMLGAATQQ AILDQITKNS LQQYGSQQEL 420
KTLIEKTNQ LLLDRIASQL SGLNPSTTGN SNNKGKKNIA TYFQLDAIGN PTLSPQKQK 480
LLLDVLDQYK DFFGTNTQAA QRDGSGKGHG SYSTYQDGS KITYLQFSYK DIDNLSLSDK 540
GNSKLASDVV AALLLFQAAD KGTQQLALSA IN 572

25 <212> Type : PRT
<211> Length : 572
SequenceName : SEQ ID 129
SequenceDescription :

30 Sequence

<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :

35 MKKFLRKPKF WLLTLGGFLS TSVILAACAT PSNSALQTVF KARSSQFFNG EQGSLQSALT 60
TALKNPVANK QFIAAPLLKA LEAWYENNED KKITQFLKDT KSNVDSQYTT AVDKVVSASR 120
NKSLFVQDDL LDNAGGSEAT WKAQKLLEQL ISDFASRVFQ KNYLNYKKDG QVSTGPFTYD 180
ELHKEESWKN MEFSAPRFSE TNDDFFAKIQ SQVFDQWVEY TDPTLISQVN YKYSAPSQGL 240
GQIYNREKLK DKLTPSYAFP FFAEEKDIAP NQNVGNKRWK QLVKGEGAIT DNNIGQSGTN 300
SQKTGLLKYR NESNKGDFLD FPLNLSDTNE TKQLVDASNI VDQLEAANLG AALNLKLQVF 360
40 EQDNDELPOI KELKEDLNNT IVVDKSKDVE KASKTNALFY NDQEGKQQQS DSDPIAGALD 420
DIFAFYFDQM KALNKKLLER ARNLNQNVNQ ANPTPWNLGL SAKLPYVNTN GNYEKLNNYF 480
TSKSKNNSAS NEVLVALKSG SINLRQVQQT DQNSYSPIKF RIVRNTGVT VFGLDGGSY 540
LKQDSTNKK VSKQSLTLT KSSSGNSNKV LRDLDKQKQF LKFRFQAKT NTFYSTNFAF 600
SFPLNETLKS WFDKRELIL ANALVNASLD QKDKASKALT EAFNPYKELI KEFAPVALAT 660
45 TMISFYFDQM KALNKKLLER ARNLNQNVNQ ANPTPWNLGL SAKLPYVNTN GNYEKLNNYF 720
TFLITKTLPV KVGQEETSI EESNKLTKT ADVDKIRDKI LENIQTKVND FVKNKLKPAL 780
APRPAYSNI LLNVNNDKVL SSGANWSLAS LLQSDKVNPL SFMLLKQAFD NNDLFKKAQK 840
LFKDIQEKSS NNGMQSSST TNSDADALSK VIGNYYTTW AKLTDKSIY NPKDNKFDEL 900
FKLAFASID EKSFNVDYKA VIDHYRFYT LQWLVDQKLE NFKSLKTNL KFGEVAFIAY 960
50 KNTETTNPSN PQGVFGSYFN YENSASEVKE STQTLDPNNF FYKTTTKPTV QAIQVASLA 1020
LVQKQMQQN STDHYGFTGL STSTSSMFD SRRDAILQOI TKTSLQYGS KDQLKKIIG 1080
TNNQLLLDRI AVQLSGLNPS TTNGSGKTI ATYFQVDAVG NPTLDFQAKR KLLLDLLDQY 1140
QNYFGNGAQ SQRDSTPSGT GNYLTYQNGS DKYTYTQFTY QDIDSLSLTT TSGTNNKIAS 1200
DVVAALLLFQ AADKGTQQLA LSAINKPQLN IGDKRIESGL KLLK 1244

55 <212> Type : PRT
<211> Length : 1244
SequenceName : SEQ ID 130
SequenceDescription :

60 Sequence

<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :

65 MVGSGAAGSA SSLQNGSNS SGLKSLLRSA PVSVPSSSTS NQTLSSLNPA PVGPQAVVSQ 60
PAGGATAAVS VNRITASDTAT FSKYLNATA LHQMGVIVPG LEKGGNGGT GVVASRRDAT 120
STNLPHAGA SQTGLGTGSP REPALTATSQ RAVTVVAGPL RAGNSSETDA LPNVITQLYH 180
TSTAQLAYLN GQIVVMSSAR VPSLWYWVG EDQESGKATW WAKTELNWGT DKQKQFVENQ 240

	LGFKDDNSND SKNSNLKTQG LTQPAYLIAG LDVVADHLVF AAFKAGAVGY DMTTDSNAST	300
	YNQALVWSTT AGLDSGGTR LW	322
	<212> Type : PRT	
	<211> Length : 322	
5	SequenceName : SEQ ID 131	
	SequenceDescription :	
	Sequence	

10	<213> OrganismName : Mycoplasma pneumoniae	
	<400> PreSequenceString :	
	MPVFLKLTHT IRKVLRLVARL SRLALLSLTA VIFSGCANIN LISAVGSSSV QPLLKSLSSH	60
	YVLNHNKDKN LVEISVQAGG SSAGVKAITK GLADIGNVSK NTKSYAEENK QLWMDKKLKT	120
	ITLKGDAIAV IYKAPSEFKG KLVLTCKDNLN DLYDLFAGSK SVDINKFVEN GQTTKNSNHN	180
15	LIGFPRTGGA FASGTAEAFI KFSGLTQTKT LDKDSKEILE GQRNYGPNAR PTSETNIEAF	240
	NTFVTTLRQP NLYGMVYLSL GFVNMMNMLI KSEGFEVLKV KYDNNAVTPS SQAVSSNTYK	300
	WVRPLNSVVS LLPKQKTLPS IQRFFNWLLF SNNSEIKKIY DDFGVLELTA DEKKKMFKTG	360
	NAEMSNIANF WVDDYSLNNQ TFGAL	385
	<212> Type : PRT	
20	<211> Length : 385	
	SequenceName : SEQ ID 132	
	SequenceDescription :	
	Sequence	

25	<213> OrganismName : Mycobacterium tuberculosis H37Rv	
	<400> PreSequenceString :	
	MSFAVLPPPEI NSARLYVGAG LAPMLDAAAA WDGLADELGS AAASFSAVTA GLAGSSWLGA	60
	ASTAMTGAAA PYLGWLSAAA AQAQQAATQT RLAAAAFEAA LAATVHPAII SANRALFVSL	120
30	VVSNLLGQNA PAIAATEAAY EQMWAQDVAA MFGYHAGASA AVSALTPFGQ ALPTVAGGGA	180
	LVSAAAAQVT TRVFRNLGLA NVGEGNVGNG NVGNFNLGSA NIGNGNIGSG NIGSSNIGFG	240
	NVGPGLTAAL NNIGFGNTGS NNIGFGNTGS NNIGFGNTGD GNRGIGLTGS GLLGFGGLNS	300
	GTGNIGLFNS GTGNVGIENS GTGNWIGENS GNSYNTGFGN SGDANTGFFN SGIANTGVGN	360
	AGNYNTGSYN PGNSNTGGFN MGQYNTGYLN SGNYNITGLN SGNVNTGAFI TGNFNNGFLW	420
35	RGDHQGLIFG SPGFFNSTSA PSSGFFNSGA GSASGFLNSG ANNSGFFNSS SGAIGNSGLA	480
	NAGVLVSGVI NSGNTVSGLF NMSLVAITTP ALISGFFNTG SNMSGFFGGP PVFNLGLANR	540
	GVNINILGNAN IGINYNILGSG NVGDFNIGLS GNLGSGNIGS SGNVGSFNIG SGNIGVFNVG	600
	SGSLGNYNIG SGNLGIYNIG FGNVGDYNVG FGNAGDFNQG FANTGNNNIG FANTGNNNIG	660
	IGLSGDNQQG FNIASGWNSG TGNSGLFNISG TNNVGIENAG TGNVGIANSI TGNWIGINPG	720
40	TDNTGILNAG SYNTGILNAG DFNTGFYNTG SYNTGGFNVG NTNTGNFNVG DTNTGSYNPG	780
	DTNTGFFNPG NVNTGAFDTG DFNNGFLVAG DNQGGIAIDL SVTTPFIPIN EQMVIDVHNV	840
	MTFPGGNMITV TEASTVFPQT FYLSGLFFFG PVNLSASTLT VPTITLTIGG PTVTVPIISIV	900
	GALESRTITF LKIDPAPGIG NSTTNPSSGF FNSGTGTSF FQNVGGGSSG VWNISGLSSAI	960
	GNSGFQNLGS LQSGWANLGN SVSGFFNTST VNLSTPANVS GLNNIGTNLS GVFRGPTGTI	1020
45	FNAGLANLGS NVFGGNQGSY NIGPANLGN NIGLGNLGSY NFGFGNAGDF NLGFANTGNN	1080
	YNIGFGNAGD FNQGFANTGN NNIGFANTGN NNIGIGLSGD NQGGFNFAAG WNSGTANIGL	1140
	FNSGTNNVGI GNSGTGNWGI GNSGSGNTGI GNTGSTNTGF FNTGIVNTGV ANAGSYNTGW	1200
	YNTGDTNTGI ANLGDFNTGF YNTGNFSTGF ANQEDIATGA FITGDMGNGA FWRGDQQGLF	1260
	SAGYRVHVPE IPAHVTVEVP VNIPITASFT NTVYSGITLE QINFGFTIDI AGIPLLAGAI	1320
50	SKAVLPPIITG TGPAITVNIG DPGGSTAIRI PATASVGPFD VTFVNIAATT GFFNATTDPS	1380
	SGFFNGGPGT VSGIANIGAN ISGFQNVANS ATSGFNNGVS LQSGLANLGD TVSGVFNTGI	1440
	GAPANVSGMF NIGSNLAGFF HDQATGMSMF NLGLGNIGQF NVGFSNVGDS NAGLANIGSF	1500
	NLGSNGLGSF NVFGGNQGSY NIGPANLGN NIGLGNLGSY NFGFGNAGDF NLGFANTGNN	1560
	NIGFANTGNN NIGIGLSGDN QQGFNFAGGW NSGSGNSGLF NSGTNNIGLF NSGTGNIGIG	1620
55	NSGTGNWGIA NTGDTNTGIF NTGDVNTGGL NAGNVNTGIF NTGHYNTGSF NAGSFNTAGF	1680
	NPGSYNTGYL NTGSYNTGLA NSGDVNTGGF ITGNYSNGFW WRGDYQGLAG ISQITITVPDT	1740
	AVPVKLHVPI FLDIPVTGTL GTFTVHGFRF PEITGDIPLI GIPFNAATLD AFSFPNISIV	1800
	LPNIGINLGS GPDPIDIDAG TGGLLPIKIP LIDIPAAPGF GNSTTPSSG FFNAGTGTVS	1860
	GVGNVGSNSS GFFNLTSGSS GISGVQNFGE LISGGFNFCN TVSGLVNAST LGLSMPANLS	1920
60	GGGNVGATVA GFVNNTQILN LGFGNVGSGN VGHGNIGDSN VGLGNLGNAN VGHGNIGSFN	1980
	VFGSNRGSYN IGPANLGNYN IGLGNLGSYN LGFANSGSNN IGFANTGNN	2040
	IGIGLSGHNQ QGFGSWNSGT ANTGLFNSGT NNIGLFNSGT GNIGIGNSGI GNTGIGNPGV	2100
	GNTGLGNSGT GNWGLWNPOT GNMGVANVGT YNTGGYNVGS TNTGIANVGI ANTGSYNTGS	2160
	TNTGSFNDGD FNTGFYNTGD YNTGFYNTGD VNTGAFIIGN FSNGAFWQSD HQGQWGAHYA	2220
65	ITVPQIPLLN IGLNIPVNL IHLDFGTLAV NGFQIPATL RALGVTHFSV GPIIVPRIAG	2280
	TLPVIDINIG DPGGSSSIPI TITSGAGPVV IPLLIDIPAP GFGNSTTGPS SGFFNSGTGS	2340
	SSGFGNVGAN NSGFWNATAF GIGNSGLQNF GSLQSGWANL GNTVSGFYNT SAADFATPAN	2400

	LSGLSNVGD	LTGVLRGPN	STFNAGLANL	GQFNVGSANL	GSANLGSANL	GSANLGSNSV	2460
	GFGNIGNANI	GGANIGDFNV	GIANTGPGLT	AAVNNIGIGN	TGNYNIGVGN	TGNYNIGFGN	2520
	TGNNNIGIGL	SGDNQIGFGP	LNAGIANMGL	FNLGDNNGFM	ANAGNFNQGI	ANTGNNNIGL	2580
	FNTGNNNVGI	WLTGDLGSGF	SSLNSGAGNT	GFFNSGTANT	GLFNSGTGNT	GLFNSGTGNV	2640
5	GIGNMTGGF	GVGLSGDSQV	GIGGTNSGSF	NIGLFNSGTG	NVGIGNSGTG	NVGIGNTGTG	2700
	NTGIGNSGNY	NTGLLNAGLV	NTGIANPGNH	NTGLFNIGTF	NTGIANPGHY	NTGSYNTGSY	2760
	NTGMANAGDY	GTGAFITGSM	NNGLLWRADR	QGLLAANYTI	TIERPAAFLN	VDIPVNIPIT	2820
	GDITNVSIPA	ITFPRIDASG	SVDIGILSGT	VLAPVGPITL	HGGDASAPLD	TPIEIDFGPS	2880
	PAINLNICKP	DGSTVINIVG	GAGAGPISIP	IIDLRPAPGF	FNATTGPSSG	FLNHWAGSAS	2940
10	GLLNFGNNSY	LYNFATSSMG	NSGFQNYGSL	QSGWANLGNS	ISGIYNTGLG	APANVSGLLN	3000
	IGTNLAGWLQ	NGPTETTFSV	GLANLGFWNL	GSANIGNYNL	GSANIGVYNL	GSANIGDFNL	3060
	GSANIGDFNL	GSANIGSSNI	GFCNVGPGLT	AAIGNIGFGN	TGNGNIGIGN	TGTGNIGFGN	3120
	TGNGNIGIGL	TGDTMTGFGG	WNSGTGNIGL	FNSGTGNIGF	GNSGTGNWGI	GNSGDYNTGI	3180
	GNTGSTNSGF	FNTGLVNTGI	GNSGDYNTGL	FNAGNTNTGS	FNPGDYNTGG	FNPGNYNTGY	3240
15	FNPGNSNTGI	ANSGDVNTGA	FNSGNYNSGF	FWRGDYQGLG	GFAYQSAVSE	IPWSYDRFQH	3300

<212> Type : PRT

<211> Length : 3300

SequenceName : SEQ ID 133

20 SequenceDescription :

Sequence

	<213> OrganismName : Mycobacterium tuberculosis H37Rv	
25	<400> PreSequenceString :	
	MNLVSTTSGM SGFLNVGALG SGVANVGNTI SGIYNVGTSD LSTPAVNSGL ANIGTNIAGL	60
	LRDGAGTAAI NLGLANHGNI NVGFASLGGF NFGCATIGHN NVGIGNTGIF DVGLANLGSY	120
	NIGFGNLGDD NLGFGNFGSY NIGFCNVGND NLGFANAGGG NIGFANTGSN NVGFGNTGSN	180
	NVGIGLTGNG QIGFGSFNSG SGNIGLFNSG SNNIGFFNSG SGNFGIANSN SFNTGIGNTG	240
30	NTNTGLFNSG DVNTGAFNPG SFNTGSFNTG SFNTGGFNPG NTNTGYLNIG NYNTGIANTG	300
	DVDTGAFITG NYSNGLFLSG DYQGLVGLNL VIDMPLPISL GVNIPIDIPI TASAGNITLM	360
	GVTIPPTGDI VLSSIAGQRA HFGPITIPNI TVVGPTTTVA IGGPNTAITI TGGGAIRIPL	420
	ISIPAAPGFG VSTTNPSSGF FNTGAGGASG FGNFGGANSN FWNLASATSG ASGLLNVGAL	480
	GSGLANVGTT NSGFYNTSTS DLATPAFNNG LANISTSIAG LLRDSTGTMV LNLGLANHGT	540
35	LNVGIANLGD YNIGFANLGS ANFGSANIGG NNIGGANTGI FDIGLANLGS YNIGFGNFGD	600
	DNLGFGNLGS YNVGFGNLGN DNLGFANTGS NNIGFANTGS NNIGIGLTGD GQIGFGSLNS	660
	GSNGIGLFNS GSGNIGFFNS GNGNVGIGNT GTANFGGLNT GSTNTGFFNS GDVNTGIGNT	720
	GSFNTGSFNP GDSNTGDFNP GSYNTGLGNT GDVDTGAFIS GSYNNGFLWS GNYQGLIGLH	780
	AALAIPEIAL TFGVDIPIHI PINIDAGVVT LQGFSSIVAAE NNIDFTPIII PTINITLPTA	840
40	AITVGGPTTS IGITASAGIG SITIPIDIPI ATSGFGNSTT SPSSGFFNSG AGSASGFLNV	900
	VAGASGISGY LNVGALGSGV TNVGHVSYGF VASALDLVT PAFASGLMRD GMGMTLNLG	960
	LANLGSNNAG FNGTGIFDVG VANLGNYNIG FGNFGDDNLG FANLGSYNIG VANTGSNNIG	1020
	FANTGSNNIG IGLTGTGQIG ICAALSGSGN IGLFNSGDGN IGFNSGTGN FGIANTGTGN	1080
	FGIGNSGSTS TGLFNSGDGN TGGFNPGNFN TGNFNTGSFN TGGFNAGNTN TGHFNTGNYN	1140
45	TGIANTGDVS TGAFISGNYS NGILWRGDYQ GLIGYSYALT IPEIPALHDV NIPIDIPITG	1200
	SFTDLVVDFN TIPIIGFESF AFSFHIHTEP DIGPIIVPSF VLSVPTFAIA VGGPTTAINI	1260
	SATAGLGPII IPIIDIPAAP GIGNSTTSPS SGFFNTGAGT ASGFGNVGGN TSGLWNLASA	1320
	ASGVSGLLNV GALGSGVANV GNTISGIYNT SPLDLGTPAF GSGLANIAGL LQGGAGTTIL	1380
	DLAGLGNLNV GLANLGGSNF GIGNTGIFNV GFANVGNHNI GLANLGNYSV GFANSGNYHI	1440
50	GIANTGSANI GFANTGSGNI GIGLTGTGQI GFGSFNSGSH NIGLFNSGDG NVGFFNSGTG	1500
	NVGIGNTGTA NFGIANSNSF NTGLGNTGST NTGLFNPGNV NTGVGNTGSI NTGSINTGSF	1560
	NTGSTNTGSF NLGDHNTGSF NSGDYNTGYF NAGDYNTGVA NTGNVNTGAF ISGNYSNGFF	1620
	WRGDYQGLIG LSTTITIPPI PYRYDLSVPI DIPITGTVVA TTPNSFTIPG FQIRVLLGPA	1680
	AVLVNEMIGP ITIDVNQVIA IDSPIQQTIS MVGTGGFGPI PIGISIGGTP GFGNSTTGPS	1740
55	SGFFHTGAGH VSGFGNFGAG NMSGSGNFGA GNSGFFNAGG LGNSGLLNFG ALQSGLANLG	1800
	NTISGVYNTS TLDLATPAFG SGIANIGANL AGLFLDNTGN LTLNFGVANQ GGLNAGIGNL	1860
	GSVNIGFVNT GDSNLGIGNL GDLNFGGVNI GGNIGIANT GIFDIGLANL GSYNIGLANL	1920
	GDDNLGFGNA GSYNIGFANF GSDNLGFANT GSYNIGFANT GNNNIGVGLT GNGQIGIGSL	1980
	NSGSMNIGLF NSGSGNIGFF NSGTGNVGIF NTGTGNFGLA NSGGFNTGIG NAGSTNTGVF	2040
60	NPGLDNTGSF NPGSFNTGGF NPGSGNTGYL NTGDYNTGVA NTGDVDTGAF ITGSYSNGFL	2100
	VSGDYQGLIG LPLLGIPTVP GYFNLTGGPS SGFFNSGAGS VSGFVNSGAG LSGYLNLTAL	2160
	GSGVANVGNT ISGVLNASAL DLATPGFLSG IGFNGTNLAG FFRG	2204

<212> Type : PRT

<211> Length : 2204

65 SequenceName : SEQ ID 134

SequenceDescription :

Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv

<400> PreSequenceString :

5	MSFVLIAP	EF	VTAAAGDLTN	LGSSISAANA	SAASATTQVL	AAGADEVSAR	IAALFGGFGL	60
	EYQAISAQVA	AYHQR	FVQAL	STGAGAYASA	EAAAAEQIVL	GVINAPTQAL	LGRPLIGDGA	120
	NATTPGGAGG	AGGLL	FGNCG	AGAAGAPGQA	GGPGGPAGLW	GNGGPGGAGG	SGGGTGGAGG	180
	AGGWLF	GVGG	AGGVGAGGG	TGGAGGPGGL	IWGGGGAGGV	GGAGGGTGGA	GGRAELLFGA	240
	GGAGGAGTDG	GPAT	TGGTGG	HGGVGGDGGW	LAPGGAGGAG	GQGGAGGAGS	DGGALGGTGG	300
10	TGGTGGAGGA	GGRGALLLGA	GGQGG	LGGAG	QGGGTGGAGG	DGVLGGVGGT	GGKGGVGGVA	360
	GLGGAGGAAG	QLFSAGGAAG	AVGVGGTGGQ	GGAGGAGAAG	ADAPASTGLT	GGTGFAGGAG		420
	GVGGQGGNAI	AGGINGS	GGGA	GGTGGQGGAG	GMGSGADNA	SGIGADGGAG	GTGGNAGAGG	480
	AGGAAGTGGT	GGVGAAGKA	GIGGTGGQGG	AGGAGSAGTD	ATATGATGTT	GFSGGAGGAG		540
	GAGGNTGVGG	TNGSGGQGGT	GGAGCAGGAG	GVGADNPTGI	GGTGGTGGKG	GAGGAGGQGG		600
15	SSGAGGTNGS	GGAGGTGGQG	GAGCAGGAGA	DNPTGIGGAG	GTGGTGGAAAG	AGGAGGAI	GT	660
	GGTGGAVGSV	GNAGTGGTGG	TGGVGGAGGA	GAAAAAGSSA	TGGAGFAGGA	GGEAGGAGNS		720
	GVGGTNGSGG	AGGAGGKG	GGT	GGAGGSGADN	PTGAGFAGGA	GGTGGAAAGAG	GAGGATGTGG	780
	TGGVVGATGS	AGTGGAGGRG	GDGGDGASGL	GLGLSGFDGG	QGGQGGAGGS	AGAGGNGAG		840
	GAGGNGGDDG	DGATGAAGLG	DNGGVGGDGG	AGGAAGNNGN	AGVGLTAKAG	DGGAAGNNGN		900
20	GGAGGAGGAG	DNNFNGGQGG	AGGQGGQGG	GGASTTSINA	NGGAGGNGGT	GGKGGAGGAG		960
	TLGVGGSGGT	GGDGGDAGSG	GGGGFAGGAAG	KAGGGGNGGR	GGDGGDGASG	LGLGLSGFDG		1020
	QGGQGGAGG	SAGAGGINGA	GGAGGNGDGG	GDGATGAAGL	GDNGGVGGDG	GAGGAAGNNG		1080
	NAGVGLTAKA	GDGGAAGNNG	NGGAGGAGGA	GDNNFNGGQGG	GAGGQGGQGG	LGGASTTSIN		1140
	ANGGAGGNGG	TGGKGGAGGA	GTGLVGGSGG	TGGDGGDAGS	GGGGGFGGAA	GKAGGGGNGG		1200
25	VGGVGGEGAS	GLGLGLSGFD	GGQGGQGGAG	GSAGAGGING	AGGAGGTGGA	GGDGAPATLI		1260
	GGPDGGDGGQ	GGIGGDDGNA	GFGAGVPDGG	GDGGNAGFGA	GVPDGGGIGG	TGGAGGAGGA		1320
	GADGDP	PSIDG	QGGAGGHGG	QGGKGG	LNST	GLASA	SGDG	1380
	GSGGTGGTGG	DAGVGG	LANT	GGTAGNAGIG	GAGGRGGDGG	AGDSCALSQD	GNGFAGGQGG	1440
	QGGVGGNAGA	GGINGAGGTG	GTGGAGGDGG	NGTTGVASEG	GAGGQGGDGG	QGGIGGAGGN		1500
30	AGFGAGVPGD	GGIGGTGGAG	GAGGAGADGD	PSIDGGQGGG	GHHGGQGGKG	GLNSTGLASA		1560
	ASGDGGNGGA	GGAGGNGDGG	DGFTGGSGGT	GGTGGDAGVG	GLANTGGTAG	NAGIGGAGGR		1620
	GGDGGAGDSG	ALSQDNGNFA	GGQGGQGGVG	GNAGAGGING	AGGTGGTGGGA	GGDQNGTTG		1680
	VASEGGAGGQ	GGDGGQGGIG	GAGGNAGFGA	GVPGDGGTGG	TGGAGGAGGA	GADGDP	PSIDG	1740
	QGGAGGHGG	QGGKGG	LNST	GLASA	SGDG			1800
35	GLGGGGGNGG	AGGAGGTPTG	SGTEGTGGDG	GDAGAGGNGG	SATGVGNGGN	GGDGGNGDGG		1860
	GNGAPGGFGG	GAGAGGLGGS	GAGGGTGDGD	GNGGSPGTGG	S			1901

<212> Type : PRT

<211> Length : 1901

SequenceName : SEQ ID 135

SequenceDescription :

Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv

<400> PreSequenceString :

45	MSLVIVAPET	VAAAAALDVAR	IGSSIGAANA	AAAGSTTSVL	AAGADEVSAA	IATLFGSHAR	60
	EYQAISTQVA	AFHDFRFAQTL	SAAVGSYVSA	EATNAAPLAT	LEHNVNLNALN	APTQALLGRP	120
	LIGDGAAGAP	GTGQAGGAGG	ILWNGGAGG	SGAPGQVGGG	GGAAGLFGTG	GAGGAGGAGA	180
	AGGAGGSGGW	LLGNGGVGGA	GGQSLGGAT	GGAGGNAGLF	GVGGTGGPGG	PGGPGGVGGT	240
50	GCAGGLGCTL	YGAGGHGGAG	GPPIGGVGG	HGGVGGAAAGL	LVGGHGGAG	GHAEGVAGA	300
	AGEDLSPHGT	SGGVGGDAGD	GGTGGRGWL	AGAGGAGGAG	GVGGTGGAGG	AGFSRALIVA	360
	GDNGGDPGAG	GAGGTGGAGS	TIGAHGAAGA	SPTSGGNGGA	GGNGAHFSSG	GKAGGNGGAG	420
	GAGGLVNGG	AGGAGNGAP	GAPPSGGDPN	GGGGAGGAG	GKGGDGGQA	GDGGAGGAGG	480
	KGGNGGNGAT	GATGLNGLGA	GADGTDGGKG	GNGGAGGGGG	AGGQGGKALA	ATHQDGSMA	540
55	GGAGCNGGAG	GMGGDGGNGA	KGTFDNGGDG	VGGNGGNGGS	RGIGGAGGIG	GAGSTAGADG	600
	ARGATPTSGG	NGGTGGNGAN	ATVAGGAGGA	GGKGGNGGLV	GNGGAGGKGG	DGMAGVAGSS	660
	PTTAGESGTS	QNGGAGGAG	GAGGRGGDFG	GDGGTGGAGG	NGANGANATT	PGAKGGDGGH	720
	GGPGAQGGNG	GQGGPGGLAG	NLFGQNGIQG	VGGSGGKGA	GGLAGDGGNG	ANGNFAGDGG	780
	NGGHGNGGN	PGAGGQGGSG	GAGSTPGAKG	AHGFTPTSGG	DGGDGGNGGN	SQVVGNGNGD	840
60	GGNGGNGGSA	GTGGNGGRGG	DGAFGGMSAN	ATNPGENGPN	GNPGGNGGAG	GAGGAGLNGG	900
	NGGAGGNGGL	GGFGNGAAG	ANGVAVGAPG	QPGGAGGHGG	AGGNGGAGGN	GGQGVVSDGA	960
	GGAGGAGDGG	GAPGDGANGG	NGQAGAFAG	GGGGRGGDGG	NAGNAGAGGP	GGTGSTAGKA	1020
	GPAGSILHDG	GNGGHGGHGA	ASGNGGGPGG	HGGNGGNGGT	GANGGNGGIG	GTGGAGSTGA	1080
	KGVLTGNEGD	GGDGGRGNG	GRGNGGGQGL	TCAGGNGGTG	GTPGNGGNGG	NGASGDLVTS	1140
65	PGDGGGGGRG	GADAGGGDAG	LGGSSGPGGT	PGDWGTGGTG	GTGGTGGQGA	NGGLTGRRG	1200
	TGGNGGNGNT	GGTGGAGGTG	GTGNGSGPG	MGGNGGAGGF	GGNGFAGVGG	RGGMGGSGGT	1260
	GCTGDAGPFG	TGTGGTGGHG	GQGGGGGFSI	LLGLGLGLGL	GSPGSIATGT	AGGAGGGGGF	1320

GGLGGGEFV 1329
<212> Type : PRT
<211> Length : 1329
SequenceName : SEQ ID 136
SequenceDescription :

5 Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
10 MSYVIATPEM MATAAFDLAR IGSQVSAASA VAAMPTTEVV AAGADEVSAG IAALFSAHAQ 60
EYQALSAQAA AFHDQFVHTL TAAARWYTAT EIANAAAMRV VLGAVNAPTQ TLLGRPLIGD 120
GAHGATAPGQP GGAGGLLFGN GGNGAAGAVG QVGGAGGAAG LFGIGGAGGA GGAGAPGGTG 180
GTGGWLAGGG GVGGMGAGG GAGGAGGNAG LFGNGGAGGA GGAGGGAGGA GGNAGWFGHG 240
15 GAGGVGGVGA AGANGATPGQ DGAAGVAGSD DGAGGDGLAG SDGGDGGAGG VGGNGGRGGW 300
LLGNGGAGGV GGVGGAGGAG AAGGAGGAGA TGINGPAGIS AAGGDGGAGG NGGAGGNGGV 360
GGAGGAGGSA GLLGYVGRAG DGGAGGGGGL GGA PGDGGAG GNGGSWLAAG DGGAGGHGGD 420
PGLGGAGGAG GASGGAGARA GANGLAAGND GPVSGGNGGK GNGGAHAPVA GGHGNGGGAG 480
GNGGLVGDGG AGGHGGDGAA GAGYADMTAI FLGSSGTPGE DGGNGGAGGA GGAGGAHAGD 540
20 GGAGGAGGNG GAGGAGGNGA HGFNAVLVSD GNGGDDGGAG GRGGDGGAGG AGGDAPAGRA 600
GSQGVGGDGG AGGAGGAPGN GSGGGRGDMA FKDDGGAGG DGGDPGAGGK GGAGGAGATE 660
GVTGATGATV HSGGNGGKGG NGADATVAGA NGCKGGAGGN GGLVGDGGAG GDGGSGAAGA 720
NGANVGEDGA DGTLSGQPGG GSEANGGQGG VGGGGAGGAG GDGGAGSSAL GSGGNGGRGD 780
AQQAGGAGGA GGAGGAGGSV SGDGGPGGKG GAGGAGGAGA SGGGGKKGAS GADSAEAVGG 840
25 AGGKGDDGGV GGVGGDGGPG GDGGAGGAAP AGQVSGHVG GVGGDGLGG AGGNGDGGH 900
GSDGGDGGDG GDFGAGGLGG LGGDSGNGTR AASGVDAADH GPGSGGNGGN GNGGAQASVA 960
GGAGGNGGDD GNAGRVGDGG AGGNGGDGAA GAGGANSAGP GSDALALGQP GNGGQGDAG 1020
QAGGAGGAGG AGGAGGSVSG DGGAGGNGGA GNGGAGGAGS GAGARGANGI DSIGGTGGAG 1080
GGGDDGGAGG VGGHGGDGGV GGAAPSGTVG SHGTGGVGGD GGLGGAGGVG GAGGNGGIGI 1140
30 TVGGAGGAGG NGGDPGAGGR GGLGGDSGNG TSAANGVDAS KHGPLTGGDG GVGNGAKAA 1200
AAGGDGGQGG DGGNAGLFGD GGAGGDGADG TAAEALGGDG GAGGAGGKGG DAGDIGDGGD 1260
GGKGGDGAHG ALGGLTVAGG NGGAGGAGGA GGAGGAFLGD GNGGAGGQGG GAGRGSGPGG 1320
GGGVGGHGA GGDAGMNGGG GTGGQGGNGA AGGAGWSPDS DLKGFDFDGG GSGGAGDGG 1380
AGGAGGTQTG DGGDGGAGGL GGAGGVGGNG VDFDINETT GRDGGDGGDG GYGWGGAGG 1440
35 NGGAGGSAPA GEVGNRGVGG DGGDGGSGGD AGNGGLGGDG FTYLADFDGE PGGDGGDGGD 1500
GGWRPQQGG GFGSTSGAHG KAGFGAPGGD GGDGNGGHHG GDGNGSFADA GDGPGGNGG 1560
NGLGGGAGRD GGAPGGDGGD GGTGSGGGFG APPRISIGG DGGDGGRGD GGRGAGGLTS 1620
GGVSSGESG GSGNGRGDPG SGGSGEGGE GGPISVNVN 1660
<212> Type : PRT
<211> Length : 1660
SequenceName : SEQ ID 137
SequenceDescription :

40 Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
45 MSFVLVSPET VAAVATDLKR IGASLAHENA SAAASTTAVV SAAADEVSTA VAALFSQHAQ 60
GYQAAAQVA AFHSRFVQAL TAGAGAYafa EAANASPLQS AMGAVSASAQ TLLSRPLIGN 120
50 GANATTTPGGN GGDGGWLFGS GGNGAPGAAG QSGGNGGSAG LWGNGGAGGA GSGGAAGGN 180
GGNGGWLFGA GGTGGIGGTG APGAMGGTGG NGGNGALLIG GGGLGGAGGM GGTGGGTGGT 240
GGNGGNGALL IGAGGVGGAG GIGGQGTGAG GAAGAGGTGG NGGAGGLFMN GGDGGAGGQG 300
GDGAAGDAAA SAGGTGGKGG QGGDGGTGGA GGAGPVLFGH GGAGGMGGQG GTGGMGGAGG 360
DGTTVIAAGT GGEAGTGGAA GAGGAAGARG ALTSGGLAGG VGAGGTGGTG GTGNGADAA 420
55 AVVGFANGD PGFAGGKGN GGIGGAATG GVAGDGGTGG KGGTGGAGGA GNDAGSTGNP 480
GGKGGDGGIG GAGGAGGAAG TGNGGHAGNT GDGGDGGTGG NGNGTGGVN GADNTLNPDT 540
PGGAGEPGGA GGAGGAGGAA GCPGGTGGTG GNGGNGGNGG NGGNGGNGN GGNAGNNSTN 600
APVGGEGGAG GDGGAGGAG AANGTAGSQ GTGGVGGDGG AGGNGGKKA GTGNSGNFGV 660
DGEAGFSGGA GNGGCVGGA GANGTGGS GNGGDDGAGG IGGAGNGIP GTGTEPAGGT 720
60 GAKGGDGGDG GAGGAGGAG GAGGQGGNAG QGGAGGAGGN AVIPGDGVGK APHGAGGSG 780
GDGKGGQGG SGGTGGSGAP IGGAGGTGG SGHAGKGA GGIGAQTII TVPNGGNGAG 840
DCGNGGAGA GNGGSGDFG GNTTSGASGS GNGGNGAGTA GSGGAGGTGG TGLSGGNGN 900
GGNGGNGDG GNGAHGTGVA QFVPATSLPT PNGGAGGNGG TGSNGGAPGP AGAPGPTTGG 960
NAGSQIGGD GNGGDDGGK GDGADAVNVV FMPTEPQAAT GTAGSAGDPT GNGGPGPTPG 1020
65 SPMVAPPPT PITQVQGGD GGAGGTGSTN ANDGTATGK GEGGVGSIL GPGGNGGTG 1080
GNASATGTNG VANAGGKGD GGGQFGAGG NGGAGGSVTD GSAGSTAGNG GNGGNATNGT 1140
IAGQPAGGNG SAGGKGGDGG NIAAGATGTA GNGGNGGNGN DGAVNAGTGG SGGNGGAGG 1200

GGANGGDGGA GGAGGAGGRG KKGIDGGFGG DGGNGGSSNG TGAGGNGGNG GTGGVGSVGA 1260
AGGDGNGGGT GGFAGFGGTA GNGGSGGTGG AGGDGGTGGD GGNGVIAGGG GTGGNGGASG 1320
AGGAGGTGGF AGNGNAGGNG GTGGASEDGD NGNAGSGATG GTGGNGGTGG DGGAAAGLGGV 1380
A 1381

5 <212> Type : PRT
<211> Length : 1381
SequenceName : SEQ ID 138
SequenceDescription :

10 Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSFVIATPEM LTTAATDLAK IGSTITAANT AAAAVAKVLP ASADEVSVAV AALFGTHAQE 60
15 YQTVSAQVAT FHDRFVQTLA AAASSYVAAE AVNVEQSLLA AVNAPTQALF GRPLIGNGAD 120
GSPGTGQAGG PGGILYGNNG NGGSGAPGQR GGAGGAAGLI GNGGNGGAGG VGTGAGAGGH 180
GGAGGWLYGN GGAGGFGGAG AVGGNGGAGG TAGLFGVGGG GGAGGNGIAG VTGTSASTPG 240
GSGTAGGAGG IGGNGGAGGA GGVLMGNGGN GGAGGEGGPG GAGGAGASGA HATNLGADGQ 300
AGGNGGNGGA GGTGCVGGPG GCHGLLGLGG SHGAGCAGGS GGDGCGAPDG GNGATGTWGH 360
20 NLGAGGTGGN GGNPAGGAG GAGGASVGGG AHGANGAPGT TSTSGGNGGD GKKGADAISS 420
GQTGANGGRG GDGGQVGNNG AGGAGGRGGA GGLFGGSEAP GRPGGAGGTG GAGGNGGTQA 480
GDGGTGGAGG AGGDGSGGA GSIGFNASAP GAAGSPGGNG GNGGPGGAGG EGGAGGLALA 540
ASGQNGSQGA GGDGAGGNG GTPCNGGHGA AGALGVNNGV GGAGGHGGDP GVGAGGQGG 600
SGSTPGANGA PGNPTSGGN GNGGGRGADA TGFGQTGASG GRGGDGGVLV NGGAGGAGGN 660
25 GSKGLPGLGR LGNPLDGGT GNGGAGGSG GAWAGNGGTG GAGGTGGVGG TGGSGSDGVN 720
GSSAGADGHP GGTGGVGGTG KKGDDGGDGG AAPNGVAGSQ GPGGAGGDGG TGGVGGNGGR 780
GIDGADGATA GARGQDGGAG GAGGKGRGG TGGPGGAGPA GTGSGQAGG NGGSGGTGGD 840
PGDGGNGANG SVFTNNGIGG NGGNGGNAGP SGAGGSGGAG STFGATGSSS SIHVNGGNGG 900
NGGNGDHALS GNGAAGGNG NGGNGSLRGS GGAGGHGNG GNASRGMGGD GGTGGAGGNA 960
30 GQIGNGGAGG NGGDGGTGS DGNPAGITSG GRGGDGGVGG QGGSVAGDGA DGGRGAGGT 1020
GGTGLRGTG ATGATGTFDA GADGHGNGG TGGVGGTGGA GGGGNGGAG GKALSPGNN 1080
GSQAGGDDG AGGAGGTGGT GGDGGRGAHG TLFSSLAGTG GTGGNGGTGG TGGTGGAGGA 1140
GGTGSTLGAT GATGAAGRAG NGGVGSGGGL GSAPGPGGTG GMGGAGGTST VSAGDGGRG 1200
GFGGDLGDS SGGNGGDDGH GGDGFRTAGA GGRGGDGGKG ADPGGLFPIP GAGGKGGTGG 1260
35 TGCTAHLGLPL AIIGQSGQPG QFGSPCADGR GGAGGAGGGG GAGGSF 1306
<212> Type : PRT
<211> Length : 1306
SequenceName : SEQ ID 139
SequenceDescription :

40 Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
45 MSAAVAWDQ LAMELASAAA SFNSVTSGLV GESWLGPSSA AMAAAVAPYL GWLAAAAQA 60
QRSATQAAAL VAEFEAVRAA MVQPALVAAN RSDLVSLVFS NFFGQNPAPAI AAIEAAYEQM 120
WAIDVSVMSA YHAGASAVAS ALTPFTAPPQ NLTDLPQALA AAPAAVVTA ITSSKGVLAN 180
LSLGLANSFG GQMGAAANLGI LNLGSLNPGG NNFGGLNVGS NNVLGNTGN GNIGFGNTGN 240
GNIGFGLTGD NQQFGGWNS GTGNIGLFNS GTGNIGIGNT GTGNFGIGNS GTSYNTGIGN 300
50 TQANTGFFN AGIANTGIGN TGNNTGFSN LGSFNTGDFN TGSSNTGFFN PGNLNTGVGN 360
TGNVNTGFFN SGNYSNGFFW RGDYQGLIGF SGTLTTPAAG LDLNGLGSVG PITIPSITIP 420
EIGLGINSSG ALVGPINVP ITVPAIGLGI NSTGALVGPI NIPPITLNSI GLELSAFQVI 480
NVGSISIPAS PLAIGLFGVN PTVGSIGPGS ISIQLGTPEI PAIPPFPGF PPDYVTVSGQ 540
IGPITFLSGG YSLPAIPLGI DVGGGLGPFT VFPDGYSLPA IPLGIDVGGG LGPFTVFPDG 600
55 YSLPAIPLGI DVGGGLGPFT VFPDGYSLPA IPLGIDVGGG IGPLTTPPIT IPSIPLGIDV 660
SGSLGPINIP IETAGTPGFG NSTTTPSSGF FNSGTGTSF FGNVSGSGSG FWNIAAGNLGN 720
SGFLNVGPLT SGILNFGNTV SGLYNTSTLG LATSAPHSV GNTDSQLAGF MRNAAGGTLF 780
NFGFANDGTL NLGNANLGDY NVGSGNVGSY NFGSGNIGNG SFGFGNIGSN NFGFGNVGSN 840
NLGFANTGPG LTELHNIGF GNIGGNNGYF ANIGNGNIGF GNTGTGNIGI GLTGDNQVGF 900
60 GALNSGSGNI GFNSGNGNI GFNSGNGNV GIGNSGNYNT GLGNVGNANT GLFNTGNVNT 960
GIGNAGSYNT GSYNAGDTNT GDLNPGNANT GYLNLGDLNT GWGNIGDLNT GALISGSYSN 1020
GILNRGDYQG LIGYSDTLSI PAIPLSVEVN GGIGPIVVPD ITIPGIPLSL NALGGVGPV 1080
VPDITIPGIP LSLNALGGVG PIVVPDITIP GIPLSLNALG GVGPIVVPDI TIPGIPLSLN 1140
ALGGVGPV PDIPTIPGIP LSLNALGGVG ITVPGVPIR IPLTINIRIP VNITLNEPFP 1200
65 NVAGIFTGYI GPIPLSTFVL GVTLAGGTLE SGIQGSFVNP FGLNIPLSGA TNAVITIPGA 1260
INPFGLVNPL SGTSPVTIP GFAINPFLN VPLSGGTSPV TIPGFTIPGS PLNLTANGGL 1320
GPINIPINIT SAPGFGNSTT TPSSGFFNSG DGSASGFGNV GPGISGLWNQ VPNALQGGVS 1380

GIYNVGQLAS GVANLGNTVS GFNNTSTVGH LTAAFNSGVN NIGQMLLGFF SPGAGP

1436

<212> Type : PRT

<211> Length : 1436

5 SequenceName : SEQ ID 140
SequenceDescription :

Sequence

10 <213> OrganismName : Mycobacterium tuberculosis H37Rv

<400> PreSequenceString :

MEFPVLPPEI	NSVLMYSGAG	SSPLLAAAAA	WDGLAEELGS	AAVSFGQVTS	GLTAGVWQGA	60
AAAAMAAAAA	PYAGWLGSVA	AAAEAVAGQA	RVVVGVFEAA	LAATVDPALV	AANRRLVAL	120
AVSNLLGQNT	PAIAAAEAAY	ELMWAADVAA	MAGYHSCASA	AAAALPAFSP	PAQALGGGVG	180
15 AFLTALFASP	AKALSLNAGL	GNVGNYNVGL	GNVGVFNLGA	GNVGGQNLGF	GNAGGTNVGF	240
GNLGNNGVGF	GNSGLGAGLA	GLGNIGLGNA	GSSNYGFANL	GVGNIGFGNT	GTNNVGVGLT	300
GNHLTGIGGL	PSGTGNIGLF	NSGTGNVGF	NSGTGNVGF	NSGNVNTGVG	NAGTASTGLF	360
NAGNFNTGVV	NVGSYNTGSF	NAGDTNTGGF	NPGGVNTGWL	NTGNTNTGIA	NSGNVNTGAF	420
ISGNFNNVGL	WVG DYQGLFG	VSAGSSIPAI	PIGLVLNGDI	GPITIQPIPI	LPTIPLSIHQ	480
20 TVNLGLPLVVP	DIVIPAFGGG	IGIPINIGPL	TITPITLFAQ	QTFVNQLPFP	TFSLGKITIP	540
QIQTTFDSNGQ	LVSFIGPIVI	DTTIPGPTNP	QIDLTIKWDI	PPITLFPNGI	SAPDNPLGLL	600
VSVSISNPGF	TIPGFSVPAQ	PLPLSIDIEG	QIDGFSTPPI	TIDRIPLTVG	GGVTIGPITI	660
QGLHIPAAPG	VGN TTTAPSS	GFNSGAGGV	SGFGNVGAGS	SGWVNQAPSA	LLGAGSGVGN	720
VGTLGSGVLN	LSSGISGFYN	TSVLPFGTPA	AVSGIGNLQ	QLSGVSAAGT	TLRSMLAGNL	780
25 GLANVGNFNT	QFGVNLGVLN	GAANIGHNL	GLGNVGDGNL	GLGNIGHGNL	GFANLGLTAG	840
AAGVGNVGF	NAGINNYGLA	NMGVGNIGFA	NTGTGNIGIG	LVGDHRTGIG	GLNSGIGNIG	900
LFNSGTGNVG	FFNSGTGNFG	IGNSGRFNTG	IGNSGTASTG	LFNAGSFSTG	IANTGDYNTG	960
SFNAGDTNTG	GFNPGGINTG	WFNTGHANTG	LANAGTFGTG	AFMTGDYSNG	LLWRGGYEGE	1020
VGVRVGPTIS	QFPVTVHAIG	GVGPHVAPV	PVPAVHVEIT	DATVGLGPFT	VPPISIPSLP	1080
30 IASITGSVDL	AANTISPIRA	LDPLAGSIGL	FLEPFRISDP	FITIDAFQV	AGVLFLENII	1140
VPGLTVSGQI	LVTPTPIPLT	LNLDTTPWTL	FPNGFTIPAQ	TPVTVGMEVA	NDGFTFFPGG	1200
LTFPRASAGV	TGLSVGLDAF	TLLPDGFTLD	TVPATFDGTI	LIGDIPPIPI	DVPAVPGFGN	1260
TTTAPSSGGF	NTGGGGSGF	ANVGAGTSGW	WNQGHVDLAG	AGSGVANAGT	LSSGVLNVGS	1320
GISGWYNTST	LGAGTPAVVS	GIGNLGQQLS	GFLANGTVLN	RSPVNIWGA	DVGAFNTGLG	1380
35 NVGDLNWGAA	NIGAQNLGLG	NLGSNGVGF	NIGAGNVGFA	NSGPAVGLAG	LGNVGLSNAG	1440
SNNWGLANLG	VGNIGLANTG	TGNIGIGLVG	DYQTGIGGLN	SGSGNIGLFN	SGTGNVGFNF	1500
TGTGNVGLFN	SGSGNIGLGN	SGTGTGLFNG	AGNFNTGIAN	PGSYNTGSFN	VGDTNTGGFN	1560
PGDINTGWFN	TGIMNTGTRN	TGALMSGTDS	NGMLWRGDHE	GLFGLSYGIT	IPQFPPIRIT	1620
TGGIGPIVIP	DTTILPPLHL	QTGTDADYSF	TVPDIPPIAI	HIGINGVVTV	GFTAPEATLL	1680
40 SALKNNGSFI	SFGPITLSNI	DIPPMDFTLG	LPVLGPITGQ	LGPILHLEPI	VAGIGVPLEI	1740
EPIPLDAISL	SESIPIRIPV	DIPASVIDGI	SMSEVVPIDA	SVDIPAVTIT	GTTISAIPLG	1800
FDIRTSAGPL	NIPIIDIPAA	PGFGNSTQMP	SSGFFNTGAG	GGSGIGNLGA	GVSGLLNQAG	1860
AGSLVGLTSG	LGNAGTLASG	VLSNGTAISG	LFNVSTLDAT	TPAVISGFSN	LGDHMSGVSI	1920
DGLIAILTFP	PAESVFDQII	DAAIAELQHL	DIGNALALGN	VGGVNLGLAN	VGEFNLGAGN	1980
45 VGNINVGAGN	LGNSNLGLGN	VGTGNLGFNG	DIGAGNFGFN	AGLTAGAGGL	GNVGLGNAGS	2040
GSWGLANVGV	GNIGLANTGT	GNIGIGLTGD	YRTGIGGLNS	GTGNLGLFNS	GTGNIGFFNT	2100
GTGNFGLFNS	GSYSTGVGNA	GTASTGLFNA	GNFNTGLANA	GSYNTGSLNV	GSFNTGGVNP	2160
GTVNTGWENT	GHTNTGLFNT	GNVNTGAFNS	GSFNNGALWT	GDYHGLVGF	FSIDIAGSTL	2220
LDLNETNLNG	PIHIEQIDIP	GMSLFDVHEI	VEIGPFTIPQ	VDVPAIPLEI	HESIHMDDIV	2280
50 LVPATTIPAQ	TRTIPLDIPA	SPGSTMTLPL	ISMRFEGEDW	ILGSTAAIPN	FGDPFPAPTQ	2340
GITIHTGPGP	GTTGELKISI	PGFEIPQIAT	TRFLLDVNIS	GGLPAFTLFA	GGLTIPTNAI	2400
PLTIDASGAL	DPITIFPGGY	TIDPLPLHLA	LNLTVPDSSI	PIIDVPPTPG	FGNTTATPSS	2460
GFNSGAGGV	SGFGNVGSNL	SGWVNQAASA	LAGSGSGVLN	VGTLGSGVLN	VSGSVSGIYN	2520
TSVLPPLGTPA	VLSGLGNVGH	QLSGVSAAGT	ALNQIPILNI	GLADVGNFNV	GFGNVGDVNL	2580
55 GAANLGAQNL	GLGNVGTGNL	GFANVGHGNI	GFGNSGLTAG	AAGLGNTGFG	NAGSANYGFA	2640
NQGVNRNIGLA	NTGTGNIGIG	LVGDNLGTIG	GLNSGAGNIG	LFNSGTGNIG	FFNSGTGNFG	2700
IGNSGSFNTG	IGNSGTGSTG	LFNAGSFNTG	VANAGSYNTG	SFNAGDTNTG	GFNPGTINTG	2760
WFNTGHTNTG	IANSGNVGTG	AFMSGNFSNG	LLWRGDHEGL	FSLFYSLDVP	RITIVDAHLD	2820
GGFGPVVLEP	IPVPAVNAHL	TGNVAMGAFT	IPQIDIPALT	PNITGSAAFR	IVVGSVRIPP	2880
60 VSVVIVEQIN	ASVGAEMRID	PFEMWTQGTN	GLGITFYSGF	SADGSPYATG	PLVFGAGTSD	2940
GSHLTISASS	GAFTTPQLET	GPITLGFQVP	GSVNATILFP	GGLTFPATSL	LNLDTVAGAG	3000
GVDIPATWTP	EIAASADGSV	YVLASSIPLI	NIPPTPGIGN	STITPSSGFF	NAGAGGGSGF	3060
GNFGAGTSGW	WNQAHTALAG	AGSGFANVGT	LHSGVLNLGS	GVSGIYNTST	LGVGTPALVS	3120
GLGNVGHQLS	GLLSGGSAVN	PVTVLNIGLA	NVGSNAGFG	NVGEVNLGAA	NLGAHNLGFG	3180
65 NIGAGNLGFG	NIGHGNVGVG	NSGLTAGVPG	LGNVGLNAG	GNNWGLANVG	VGNIGLANTG	3240
TGNIGIGLTG	DYQTGIGGLN	SGAGNLGLFN	SGAGNVGFNF	TGTGNFGLFN	SGSFNTGVGN	3300
SGTGSTGLFN	AGSFNTGVAN	AGSYNTGSFN	VGDTNTGGFN	PGSINTGWLN	AGNANTGVAN	3360

	AGNVNTGAFV	TGNFSNGILW	RGDYQGLAGF	AVGYTLPLFP	AVGADVSGGI	GPITVLPPIH	3420
	IPPIPVGFAA	VGGIGPIAIP	DISVPSIHLG	LDPAVHVGS	TVNPITVRTP	PVLVSYSQGA	3480
	VTSTSGPTSE	IWKVPSFFPG	IRIAPSSGGG	ATSTQGAYFV	GPISIPSGTV	TFPGFTIPLD	3540
	PIDIGLPLVSL	TIPGFTIPGG	TLIPTLPLGL	ALSNIGPPVD	IPAIVLDRIL	LDLHADTTIG	3600
5	PINVPITAGFG	GAPGFGNSTT	LPSSGFFNTG	AGGSGGFSNT	GAGMSGLLNA	MSDPLLSAS	3660
	GFANFGTQLS	GILNRGAGIS	GVYNTGALGV	VTAAVVSFGF	NVGQQLSGLL	FTGVGP	3716

<212> Type : PRT

<211> Length : 3716

10 SequenceName : SEQ ID 141
 SequenceDescription :

Sequence

15 <213> OrganismName : Mycobacterium tuberculosis H37Rv

<400> PreSequenceString :

	MNFPVLPEPEI	NSVLMYSGAG	SSPLLAASAAA	WDGLAEELGS	AAVSFGQVTS	GLTAGVWQGA	60
	AAAAAASAAA	PYAGWLGSVA	AQAVAVAGQA	RAAVAAFEAA	LAATVDPAAV	AVNRMAMRAL	120
	AMSNLLGQNA	AAIAAVEAEY	ELMWAADVAA	MAGYHSGASA	AAAALPAFSP	PAQALGGGVG	180
20	AFLNALFAGP	AKMLRLNAGL	GNVGNYNVGL	GNVGIFNLGA	ANVGAQNLGA	ANAGSGNFGF	240
	GNIGNANFGP	GNSGLGLPPG	MGNIGLGNAG	SSNYGLANLG	VGNIGFANTG	SNNIGIGLTG	300
	DNLTGIGGLN	SGTGNLGLFN	SGTGNIGFFN	SGTGNFGVFN	SGSYNTGVGN	AGTASTGLFN	360
	VGGFNTGVAN	VGSYNTGSFN	AGNTNTGGFN	PGNVNTGWLN	TGNTNTGIAN	SGNVNTGAFI	420
	SGNFSNGVLW	RGDYEGLWGL	SGGSTIPAIP	IGLELNGGVG	PITVLPPIQL	PTIPLNIHQT	480
25	FSLGLPLVVD	IVIPAFGGGT	AIPISVGPIT	ISPTILFPAQ	NFNTTFFVGP	FFGLGVVNIS	540
	GIEIKDLAGN	VTLQLGNLNI	DTRINQSFV	TVNWSTPAVT	IFPNGISIPN	NPLALLASAS	600
	IGTLGFTIPG	FTIPAAPLPL	TIDIDGQIDG	FSTPPTITDR	IPNLNGASVT	VGPILINGVN	660
	IPATPGFGNT	TTAPSSGGFN	SGDGGVSGFG	NFGAGSSGWW	NQAQTEVAGA	GGGFANFGSL	720
	GSGLNLNFGS	GSGLYNTGGL	PPGTPAVVSG	IGNVGEQLSG	LSSAGTALNQ	SLIINLGLAD	780
30	VGSVNVGFGN	VGDFNLGAAN	IGDLNVGLGN	VGGGNVGFNG	IGDANFGLGN	AGLAAGLAGV	840
	GNIGLGNAGS	GNVGFNGMGV	GNIGFGNTGT	NNLGLGLTGD	NQTGIGGLNS	GAGNIGLFNS	900
	GTGNVGLFNS	GTGNFGLFNS	GSFNTGIGNG	GTGSTGLFNA	GNFNTGVANP	GSYNTGSFNV	960
	GDNTGFGFNP	GSIPNTGFNT	GNANTGVANS	GNVDTGALMS	GNFSNGILWR	GNFEGFLGLN	1020
	VGITIPPEFPI	HWTSTGGIGP	IIIPDPTIIP	PIHLGLTGQA	NYGFAVPDIP	IPAIHIDFDG	1080
35	AADAGFTAPA	TTLLSALGIT	GQFRFGPITV	SNVQLNPFNV	NLKLQFLHDA	FPNEFPDPTI	1140
	SVQIQVAIPL	TSATLGGLAL	PLQQTIDAIE	LPAISFSQSI	PIDIPPIDIP	ASTINGISMS	1200
	EVVPIIDVPA	IPAVTITGTR	IDPIPLNFDV	IGDIPALNIS	IIDIPALPGF	GNSTELPSSG	1260
	FFNTGGGGGS	GIANFGAGVS	GLLNQASSPM	VGTLSGLGNA	GSLASGVVLS	GVDISGMFNV	1320
	STLGSAPAVI	SGFNLGNHNV	SCVSIDGLLA	MLTSGSGSGS	GQPSIIDAAI	AELRHLNPLN	1380
40	IVNLGNVGSY	NLGFANVGDV	NLGAAGNLGN	NLGGGNLGGQ	NLGLGNLGDG	NVGFGNLGHG	1440
	NVGFNLNGLG	ALPGIGNIGL	GNAGSNNVGF	GNMGLGNIGF	GNVTGNLNLGI	GLTGDNTQGF	1500
	GGLNSGAGNL	GLFNSGTGNI	GGFNTGTGNW	GLFNSGSYNT	GIGNSGTGST	GLFNAGSFNT	1560
	GLANAGSYNT	GSLNAGNTNT	GGFNPNGVNT	GWFNAGHTNT	GGFNTGNVNT	GAFNSGSFNN	1620
	GALWTGDHGH	LVGFYSYSIE	TGSTLVVDINE	TLNLGPVHID	QIDIPGMSLF	DIHELNVNIGP	1680
45	FRIEPIDVPA	LVGDIHETMV	IPPIVFLPSM	TIGGQTYTIP	LDTPPAAPP	PFRPLPLFVN	1740
	ALGDNWIVGA	SNSTGMSGGF	VTAPTQIGILI	HTGPSSATTG	SLALTLPTVT	IPITITSPIP	1800
	LKIDVSGGLP	AFTLFPGLLN	IPQNAIPLTI	DASGVLDPIIT	IFPGGFTIDP	LPLSLALNIS	1860
	VPDSSVPIII	VPPTPGFGNA	TATPSSGGFN	SGAGGVSGFG	NFGAGSSGWW	NQAHAALAGA	1920
	GSGLNVGTL	NSGVLNVGSG	ISGLYNTAIV	GLGTPALVSG	AGNVGQQLSG	VLAAGTALTQ	1980
50	SPIINLGLAD	VGNYNLGLGN	VGDFNLGAAN	LGDLNLGLGN	IGNANVGFGN	IGHGNVGFNG	2040
	SGLGAALGIG	NIGLGNAGST	NVGLANMGVG	NIGFANTGTN	NLGLGLTGDN	QTGIGGLNSG	2100
	AGNIGLFNSG	TGNIGFFNSG	TGNWGLFNSG	SFNTGIGNSG	TGSTGLFNAG	GFTTGLANAG	2160
	SYNTGSENVG	DTNTGGFNPG	SINTGWFTNG	NANTGIANSG	NVDTGALMSG	NFSNGILWRG	2220
	NYEGLFSYSY	SLDVPRITIL	DAHFTGAFGP	VVVPPIPVLA	INAHLTGNAA	MGAFTIPQID	2280
55	IPALNPNVTG	SVGFPGPIAVP	SVTIPALTAA	RAVLDMASV	GATSEIEPFI	VWTSSGAIGP	2340
	TWYSVGRIYN	AGDLFVGGNI	ISGIPTLSTT	GPVHAVFNAA	SQAFNTPALN	IHQIPLGFQV	2400
	PGSIDAITLF	PVGLTFPANS	LLNLDFVFGT	PGATIPAITF	PEIPANADGE	LYVIAGDIPL	2460
	INIPPTPGIG	NTTVPSSSGF	FNTGAGGGSG	FGNFGANMSG	WWNQAHTALA	GAGSGIANVG	2520
	TLHSGVLNLG	SGLSGIYNTS	TLPLGTPALV	SGLGNVGDHL	SGLLASNVGQ	NPITIVNIGL	2580
60	ANVGNGNVGL	GNIGNLNLGA	ANIGDVNLGF	GNIGDVNLGF	GNIGGGNVGF	GNIGDANFGF	2640
	GNSGLAAGLA	GNIGNLNLGA	GSNGVGNWAM	GLNIGFGNT	GTNNLIGILT	GDNQSGIGGL	2700
	NSGTGNIGLF	NSGTGNIGFF	NSGTANFGLF	NSGSYNTGIG	NSGVASTGLV	NAGGFNTGVA	2760
	NAGSYNTGSF	NAGDTNTGGF	NPGSTNTGWF	NTGNANTGVA	NAGNVNTGAL	ITGNFSNGIL	2820
	WRGNYEGLAG	FSFGYPIPLF	PAVGADVTDG	IGPATIIPPI	HIPSIPLGFA	AIGHIGPISI	2880
65	PNIAIPSIHL	GIDPTFDVGP	ITVDPITLTI	PGLSLDAVVS	EIRMTSGSSS	GFKVRPSFSF	2940
	FAVGPDGMPG	GEVSILQPF	VAPINLNPTT	LHFPGFTIPT	GPIHIGLPLS	LTIPGFTIPG	3000
	GTLIPQLPLG	LGLSGGTPPF	DLPTVVIDRI	PVELHASTTI	GPVSLPIFGF	GGAPGFGNDT	3060

TAPSSGFFNT GGGGSGFSN SGSGMSGVLN AISDPLLGSA SGFANFGTQL SGILNRGAGI 3120
SGVYNTGTLG LVTSAFVSGF MNVQQQLSGL LFAGTGP 3157
<212> Type : PRT
<211> Length : 3157
5 SequenceName : SEQ ID 142
SequenceDescription :

Sequence

10 <213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSFVVMPEEI NSLLIYTGAG PGPLLAASAAA WDELAELGSG AAAAFGVSPTS GLVGGIWOQGP 60
SSVAMAAAAA PYAGWLSAAA ASAESAAQQA RAVVGVFEEA LAETVDPFVI AANRSRLVSL 120
ALSNLFGQNT PAIAAAEFDY ELMWAQDVAA MLGYHTGASA AAEALAPFGS PLASLAAAAE 180
15 PAKSLAVNLG LANVGLFNAG SGNVGSYNVG AGNVGSYNVG GGNIGGNNVG LGNVGWGNFG 240
LGNSGLTPGL MGLGNIGFGN AGSYNFGLAN MGVGNIGFAN TGSGNFGIGL TGDNLGTGFG 300
FNTGSGNVGL FNSGTGNVGF FNSGTGNWGV FNSGSYNTGI GNSGIASLGL FNAGGFNTGV 360
VNAGSYNTGS FNAGEANTGG FNPGSVNTGW LNTGDINTGV ANSGDVNTGA FISGNYSNGV 420
LWRGDYQGLL GFSSGANVLP VIPLSLDING GVGAITIEPI HILPDIPINI NETLYLGPLV 480
20 VPPINVPAS LGVGIPNISI GPIKINPITL WPAQNFNQTI TLAWPVSSIT IPQIQQVALS 540
PSPIPTTLGL PIHINTGFST PVTFSYSTPA LTLFVPGLSI PTGGPLTLTL GTAGTEAFT 600
IPGFSIPEQP LPLAINVIGH INALSTPAIT IDNIPLNLHA IGGVGPVDIV GGNVPASPGF 660
GNSTTAPSSG FFNTGAGGVS GFGNVGAHTS GWFNQSTQAM QVLPGTVSGY FNSGTLMSGI 720
GNVGTQLSGM LSGGALGGNN FGLGNIGFDN VGFGNAGSSN FGLANMGIGN IGLANTGNNG 780
25 IGIHLSGDNL TGFSGFNSGS ENVLGNISGT GNVGSFNSGT GNLGVFNSSG HNTGFFLTGN 840
NINVLAPFTP GTLFTISEIP IDLQVIGGIG PIHVQPIDIP AFDIQTGGF IGIREFTLPE 900
ITIPAIPHV TGTVGLGEGF VNPAPVLFQ TAMAETADP VVLPDPFITI DHYGPPLGPP 960
GAKFPSSGFY LSISDLQING PIIGSYGGPG TIPGPFPGATF NLSTSSIALF PAGLTVPDQT 1020
PVTVNLTGGL DSITLFPGL AFPENPVVSL TNFSVGTGGF TVFPQGFTVD RIPVDLHTTL 1080
30 SIGPFPFRWD YIPPTPANG IPAVPGGFGL TSGLFPFHFT LGGGIGPISI PTTTVVDALN 1140
PLLTVTGNLE VGPFTVPDIP IPAINFGLDG NVNVSFNAPA TTLLSGLGIT GSIDISGIQI 1200
TNIQTQPAQL FMSVGQTLFL FDFRDGIELN PIVIPGSSIP ITMAGLSIPL PTVSESIPLN 1260
PSFGSPASTV KSMILHEILP IDVSINLEDA VFIPATVLPV IPLNVDVTIP VGPINIPIT 1320
EPGSGNSTTT TSDPFSGLAV PGLGVGLLGL FDGSIANNLI SGFNSAVGIV GPNVGLSNLG 1380
35 GGNVGLGNVG DFNLGAGNVG GFNVGGGNIG GNVVGLGNVG FGNVGLANS LTPGLMGLGN 1440
IGFGNAGSYN FGLANMGVGN IGFAANTGSGN FGIGLTGDNL TGFGGFNTGS GNVGLFNSTG 1500
GNVGFNNGT FLANLFPSTP VVTIDEIPLL ASITGHSEPV DIFPGGLTIP AMNPLSINLS 1560
ANTGGFNP GS VNTGWLNTGD INTGVANS GD VNTGAFISGN YSNGAFWRGD YQGLLGFSYR 1620
PAVLPQTPEL DLTLTGGLGS VVIPAIDIPA IRPEFSANVA IDSFTVPSIP IPQIDLAATT 1680
40 VSVGLGPITV PHLDIPRPV TNLVLFSGSQ GGPLKIGPIT GLFNTPIGLT PLALSQIVIG 1740
ASSSQGTITA FLANLFPSTP VVTIDEIPLL ASITGHSEPV DIFPGGLTIP AMNPLSINLS 1800
GGTGAVTIPA ITIGEIPFDL VAHSTLGPVH ILIDLPAVPG FGNTTGAPSS GFFNSGAGGV 1860
SGFGNVGAMV SGGWNQAPSA LLGGSGGVFN AGTLHSGVLN FGSGMSGFLN TSVLGLGAPA 1920
LVSGLSVGVQ QLSGLLASGT ALHQGLVLNF GLADVGLGNV GLGNVGDFNL GAGNVGGFNV 1980
45 GGGNIGGNV GLGNVGWGNF GLGNSGLTPG LMGLGNIGF NAGSYNFGLA NMGVGNIGFA 2040
NTGSGNFGIG LTGDNLTGFG GFNTGSGNVG LFNSGTGNVG FFNSGTGNWG VFNSGSYNTG 2100
IGNSGIASL LFNAGGFNTG VVNAGSYNTG SFNAGQANTG GFNPGSVNTG WLNTGDINTG 2160
VANSGDVNTG AFISGNYSNG AFWRGDYQGL LGFSYSTSTII PEFTVANIHA SGGAGPIIVP 2220
SIQFPAIPLD LSATGHIGGF TIPPVSISPI TVRIDPVFDL GPITVQDITI PALGLDPAATG 2280
50 VTVGPIFSSG SIIDPFSLTL LGFINVNVA IQTAPSEILP FTVLLSSLGV THLTPEITIP 2340
GFHIPVDPIH VELPLSVTIG PFVSPFITIP QLPLGLALSG ATPAFAPPLE ITIDRIPVVL 2400
DVNALLGPIN AGLVIPPVPG FGNTTAVPSS GFFNIGGGGG LSGFHNLAG MSGVLNAISD 2460
PLLSASGFA NFGTQLSGIL NRGADISGVY NTGALGLITS ALVSGFGNVG QQLAGLIYTG 2520
TGP 2523
55 <212> Type : PRT
<211> Length : 2523
SequenceName : SEQ ID 143
SequenceDescription :

Sequence

60 <213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSFVIAVPEA LTMAASDLAN IGSTINAANA AAALPTTGTV AAAADEVSAA VAALFGSYAQ 60
65 SYQAFGAQLS AFHAQFVQSL TNGARSYVVA EATSAAPLQD LLGVVNAPAQ ALLGRPLIGN 120
GANGADGTGA PGGPGGLLLG NGGNGSGCAP GQPGGAGGDA GLIGNGGTGG KGGDGLVGSG 180
AAGGVGGRGG WLLNGGTGG AGGAAGATLV GGTGGVGGAT GLIGSGGFGG AGGAAAGVGT 240

5 TGGVGGSGGV GGVFGNGGFG GAGGLGAAGG VGGAASYFGT GGGGGVGGDG APGGDGGAGP 300
LLIGNGGVGG LGGAGAAAGN GGAGMLLGD GGAGGQGGPA VAGVLGGMPG AGNGGNANW 360
FGSGGAGGQG GTGLAGTNGV NPGSIANPT GANGTDNSGN GNQTGGNGGP GPAGGVGEAG 420
GVGGQGGGLGE SLGDNDGTGG KGGAGGTAGT DGGAGGAGGA GGIGETDGS A GGVATGGEGG 480
5 DGATGGVDGG VGGAGGKGGQ GHNTGVGDAF GGDGGIGGDG NGALGAAGCN GGTGGAGNG 540
GRGGMILIGN GAGGAGGTGG TGGGGAAGFA GGVGGAGGEG LTDGAGTAEG GTGGLGGLGG 600
VGGTGGMGGG GGVGGNGGAA GSLIGLGGGG GAGGVGGTGG IGGIGGAGCN GGAGGAGTTT 660
GGGATIGGGG GTGGVGGAGG TGGTGGAGGT TGGSGGAGGL IGWAGAAGGT GAGGTGGQGG 720
10 LGGQGGNGGN GGTGATGGQG GDFALGGNGG AGGAGGSPGG SSGIQNMGP PGTQGADG 778

<212> Type : PRT
<211> Length : 778
SequenceName : SEQ ID 144
SequenceDescription :

15 Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :

20 PQGADGNAGN GGDGGVGGNG GNGADNTTAA AAGTTGGAGG AGGAGGTGGT GGAAGTGTGG 60
QQNGNGNGGN GGTGGKGGTG GDGALAGSSG GAGGKGGNGG DAGKAGTGSA PGTAGTGGDG 120
GKGGNGGIGA AGTTGPVGTG ASGGTGGSGG AGGTGGDGA ANGGTAGAGG AGNGGKGGD 180
GGAGVTSSTA GNSGGAGGSG GKGGDAGAGG AGATPGANGI AGNGGDGGDG AAGAVGISGA 240
TGAGDGGHGG TGAAGGNGGT GGAGGSGIDG VGGGTGGTGG NGNGAIGGA GGDAGGSGNS 300
25 GNGNGIGGGG GNAGAGGAAG NNGGTVGANG TGGDGGNGGA AGAATAGSNG GAGTGSAGGN 360
GGTGGRGSSG GAGGDGIGGV GGGKGGNGAD GEVGGAGGAG GSGPNTSPGG NGGQGGQGGG 420
GGAGGAAGAG GAGGGANGTA GNGGQGGAGG TGGAGAASSA TNGGSGGAGG TGGDGGSGGA 480
GGTGGAGGTG GAAGDGGQGG QGGAGGGAGG QGGAGGAGGT GNGGNGITGG TAGTAGAAGN 540
GGAAGKGGAG GQGGTGGGTG QGGAGGDDG AGGTGGDRTV GGGTVPAGSG GQGGNAGGGG 600
30 AGGQGGADGG SGGDGGDAGT GNGNGNGNR NSGNGTGGAG GNGGGGANGG AGGAGGSGGG 660
TGGNGGAGGD AGDAGNGNG NGTNGNGNGG NGGIAGMGN GAGTGSNG GNGGSGNGG 720
NAGMGNSGT GSGDGGAGGN GGAAGTGGTG GDGGLTGTGG TGGSGGTGGD GNGGNGADN 780
TANMTAAGG DGGNGGDDGF GGGAGAGGGG LTAGANGTGG QGGAGDGCN GAI GGHGPLT 840
DDPGGNGGTG GNGGTGGTGG AGIGSLGGT GGDGGNGNG GTGEGGGEVG GAGGTGGAAG 900
35 NGGDGGTGGT GGGDGGAGGT GGTGGTGGLG DPRVGGSGGD GGTGGSGGAA GNGGNGNAG 960
AGGNGNGGTG GAGGIGGTGG NGGDAEPGVP PGAGGAGGAG TTGGKGGTGG NGSCTGSGGT 1020
GGDGGTGGGG GNGGTGWNNG KGD TGSGGGA GDGKAPAGG TGGAGDGGA GKGSGSGV 1079

<212> Type : PRT
<211> Length : 1079
SequenceName : SEQ ID 145
SequenceDescription :

40 Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :

45 MVMSLMVAPE LVAAAAADLT GIGQAISAN AAAAGPTTQV LAAAGDEVSA AIAALFGTHA 60
QEYQALSARV ATFHEQFVRS LTAAGSAYAT AEAANASPLQ ALEQQVLGAI NAPTQLWLGR 120
50 PLIGDGVHGA PGTGQPGGAG GLLWNGGNG GSGAAGQVGG PGGAAGLFCN GSGGSGGAG 180
AAGGVGGSGG WLNNGGAGG AGGTGANGGA GGNAWLFGAG GSGGAGTNGG VGGSGGFVYG 240
NGGAGGIGGI GGIGGNGGDA GLFNGGAGG AGAAGLPGAA GLNGGDGSDG GNGGTGGNGG 300
RGGLLVNGG AGGAGGVGGD GKGGAGDPS FAVNNGAGGN GGHGNGPGVG GAGGAGLLA 360
GAHGAAGATP TSGGNGGDDG IGATANSPLQ AGGAGGNGGH GGLVNGGTTG GAGGAGHAGS 420
55 TGATGTALQP TGGNGTNGGA GGHGNGGNG GAQHGDGGVG GKGGAGGSGG AGGNGFDAAT 480
LGSPGADGGM GNGGKGGDG GKAGDGGAGA AGDVTLAVNQ GAGGDGGNGG EVGVGGKGA 540
GGVSANPALN GSAGANGTAP TSGGNGGNG AGATPTVAGE NGGAGGNGGH GGSVNGGAG 600
GAGGNGVAGT GLALNGGNG NGGIGGNGS AAGTGGDGGK GNGGAGANG QDFSASANGA 660
NNGQCGNGGN GGIGGKGGDA FATFAKAGN GAGGNGGNGV VAGQGGAGGK GAI PAMKAT 720
60 GADGTAPTSG GDGNGNGGA SPTVAGNGG DGGKGGSGGN VNGGNGGAG GNGAAGQAGT 780
PGTSGDSTG SGTGGAGGN GGAGGAGTL AGHGGNGGK GNGGQGGIGG AGERGADGAG 840
PNANGANGEN GSGGNGGDDG GAGGNGGAGG KAQAAGYTDG ATGTGGDGCN GGDGGKAGD 900
GAGENGLNSG AMLPGGTVG NPPTGGNGN GGNAGVGGTG GKAGTGSITG LDGTDGITPN 960
GNGGNGGNG GKGGTAGNS GAAGNGGNG GGLNGGDAG NGGNGGALN QAGFFGTGGK 1020
65 GNGGNGGAG MINGGLGGFG GAGGGGAVDV AATTGGAGN GCAGGFASTG LGGPGGAGG 1080
GGAGDFASGV GGVGGAGGDG GAGGVGGFVG QGGTGGEGRT GNGGSGGDDG GGGISLGGN 1140
GLGGNGGVSE TGFGGAGNG GYGGPGGPEG NGGLGGNGGA GNGGVSTTG GDGAGGKGG 1200

NGGDGGNVGL GGDAGSGGAG GNGGIGTADG GAGGAGGAGG NGGSSKSTTT GNAGSGGAGG 1260
NGGTGLNGAG GAGGAGGNAG VAGVSFGNAV GGDGNGGNG GHGGDGTGG AGGKGNGSS 1320
GAASGSGVNV VTAGHGGNGG NGGNGGNSA GAGGQGGAGG SAGNGHGGG ATGGDGGNGG 1380
NGGNSGNSTG VAGLAGGAAG AGGNGGTSS AAGHGGSGGS GSGTGTGGAG AAGGNGGAGA 1440
5 GGGSLSTGQS GGP RRQRWCR WQRRRWLGRQ RRRRWCRWQR RCRRQRWRR CRQRRLRRQW 1500
RQRRRRCRPW LHRRRRGRQGR RWRQRRFQQR QRSRWQRR 1538
<212> Type : PRT
<211> Length : 1538
SequenceName : SEQ ID 146
10 SequenceDescription :

Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
15 <400> PreSequenceString :
MSFVV TAPPV LASAASDLGG IASMISEANA MAAVRTTALA PAAADEVSAA IAALFSSYAR 60
DYQTL SVQVT AFHVQFAQTL TNAGQLYAVV DVGNVLLKT EQQVLGVINA PTQTLVGRPL 120
IGDGT H GAPG TGQNGGAGGI LWGNGGNGGS GAPGQPGGRG GDAGLFGHGG HGGVGGPGIA 180
GAAGT AGLP G GNGANGSGSG IGGAGGAGGN GGLLFGNGGA GGQGGSGGLG GSGGTGGAGM 240
20 AAGPAGGTGG IGGIGGIGGA GGVGGHGSAL FGHGGINGDG GTGGMGGQGG AGNGWAAEG 300
ITVGI GEQGG QGGDGGAGGA GGIGGSAGGI GGSQGAGGHG GDGGQGGAGG SGGVGGGGAG 360
AGGDG GAGGI GGTGNGSIG GAAGNGNGG RGGAGGMATA GSDGNGGGG GNGGVGVGSA 420
GGAGG TGGDG GAAGAGGAPG HGYPQQPAPQ GLPIGTGGTG GEGGAGGAGG DGGQGDIGFD 480
GGRGD DGGPG GGGGAGGDGS GTFNAQANNG GDGGAGGVGG AGGTGGTGGV GADGGRGGDS 540
25 GRGDG GGNAG HGGAAQFSGR GAYGGE GSG GAGGNAGGAG TGGTAGSGGA GFGGNGADG 600
GNGGNGGNGG FGGINGTFGT NGAGGTGGLG TLLGHNNGNI GLNGATGGIG STTLTNATVP 660
LQLVNTTEPV VFISLNGGQM VPVLLDTGST GLVMSDQFLT QNFGPVI GTG TAGYAGGLTY 720
NYNTYSTTVD FGNGLLTLPT SVNVTSSSP GTLGNFLSRS GAVGVLGIGP NNGFPGTSSI 780
VTAMPGLLNN GVLIDESAGI LQFGPNTLTG GITISGAPIS TVAVQIDNGP LQQAPVMFDS 840
30 GGINGTIPSA LASLPSGGFV PAGTTISVYT SDGQTL LYSY TTTATNTPFV TSGGVMNTGH 900
VPFAQQPIYV SYSPTAIGTT TFN 923
<212> Type : PRT
<211> Length : 923
SequenceName : SEQ ID 147
35 SequenceDescription :

Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
40 <400> PreSequenceString :
MIGNGGAGGS GAPGAIGGAG GPAGLIGVGG AGGAGGDSAV AGVIGGAGGA GGAALLFGAG 60
GAGGAGGSGG SGAAGCAGGA GGAGGLFASG GSGGFGGFAS TGTGGAGGTG GAGGLFASGG 120
VGGTGGGAGS GGTGGVGGTG GAGGLFASGG AGGAGSGSGT GGAGGTGGAG GLFGAGGAGG 180
LGGQGNHTGG HGGAGGSAGL LALGDGAGG AGGAATTGTG GAGGAGGKAG LLFGSGGAGG 240
45 SGGAAAGTFD TGNSSGAGGA GKAGLLFGS GGAGGSGGAG GFANGSTGGA GGAGGGAGLI 300
GNGGNGGSGG TSVATGGAGN GGAGGAGGGA GLIGNGNGG SSGMGDAPGG TGVGGIGGLL 360
LGLDGANAPA STNPLHTAQQ QALAAVNAPI QAVTGRPLIG NGANGAPGSG APGGHGGWLF 420
GGGGTGGSGV SGGAGDGGGA GGILFGAGGA GGAGGAVTGT GATGSGSGAG GALLFGAGG 480
AGGAGSSGI GGFAAGGAGG PGAGGLFNG GGAGGAGSG VSGGAGGEGG AGGAGGLFAG 540
50 GGAGGAGGSG NNVGGAGGAG GVGGLFGAGG AGGSGGGSV AGDSGAGGNA GLLAPGLAGG 600
AGGGGGQGF TGGAGGPGD AGLLVGSGGV GGAGGFGLTT GGPAAAGGDA GLLFGSGGAG 660
GAGGSGRTDL GGAGGAGGKA GLIGNGNGG AGGAGNGGG DGGPGGAAGF LGNGGNGGNG 720
GTGTSAGSPG AGGAGGSLIG AEGLPGLLP 749
<212> Type : PRT
55 <211> Length : 749
SequenceName : SEQ ID 148
SequenceDescription :

Sequence

60 <213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSFVIAAPEA LVAVASDLAG IGSALAEANA AALAPTTALL AAGADEVSAA IAALFGAHGQ 60
AYQTVSAQAS AFHAQFVQAL TGGGGAYAAA EAA NVSAAQS TDQRLLDLIN GPTQALLGRP 120
65 LIGDGANGGP QDGGPGGLL YGNGNGGTS TTAGVAGNG GAAGLIGNG AGGGGGAGAA 180
GGNGGAGGWL YDNGGAGG GTSVIPGVAG NGGAGGSAG LWGTGGAGG GGNRSGPVN 240
VAGSAGNGG AGGAAGLFGD AGAGNGGKKG GAGGAASF SIN FTAGDGGAGG AGSGGHALL 300

WGAGGAGGNG GSGGTGGAGG STAGAGGNGG AGGGGGTGGG LFGNGGAGGH GAAAGNGLAA 360
GNGVSSSSGG GAGGTGGAGG DGGAGGAGGN ARLWGVGGAG GAGGDGGAGG AGGKGGGSLG 420
GNANGGAGGD SGRGGTGGAG GEGGAAGLLV GTGGHGGDGG AGGAAVKGGD GGAAAGTGIA 480
GAGGRGGAGG SGGSGGDGGG GAAGPAGWLF GDGGAGGNGG AAAAGGAGGQ AGGGGGNNGN 540
5 GNGGNGGNGG GNGATGGWLY GNGGAGGQCA TAGAGGAGAN GVSSTNGGGT GNGGGIGGTG 600
GSGGAGGNAG LLGVGGAGGH GASGGACDRG GAGGTGFISS DGGAGGDGGD GNGGAGGTG 660
GLLFAGGNG GPGGSGGAAD ICGNGGAGNG GGTDCNGNGG GSGGGAGSGG DGGGAGGNGA 720
WLFNGGAGG GGGKGGNGAG GGLGGGSFGL PGLNGSGGDG GDGGNGAPGG VLYGNGGAGG 780
QSSSGGIGGP GATGGAGGKG GDGGDAQLIG DGGNGGNGGA GGTGGTPGPG GPGGSGGLGG 840
10 LLFGQTGTAG VSP 853
<212> Type : PRT
<211> Length : 853
SequenceName : SEQ ID 149
SequenceDescription :
15 Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
20 MSYLVVVPPEL VAAAATDLAN IGSSISAANA AAAAPTALV AAGGDEVSA IAAALFGAHR 60
AYQALSAQAA MFHEQFVRAL AAGGNSYAVA EAATAQSVQQ DLLNLINAPT QALLGRPLIG 120
NGANGLPGTG QNGGDGGILY GNGGNGGSGG VNQAGGNGGN AGLWNGGSGG GAGGNATTAG 180
RNGFNNGAGG SGGLLWNGGG AGGAGGNGGP APLVGGVGTG GGAGGNGGGA GLFYGFGGAG 240
GNGMGGVAP STGPSMGILP AGGVGGPGGG GGASALAFGS GGVGGAGGLG GPTDGTVQGV 300
25 GFGGQGGNG QSGLLFGNA GAGGAGAAGG AGTGDTEFSG GHGGAGGDGG AVGLIGNGGA 360
GGTGSPPGAVV GNGGVGGGLG GAGSPGGLLY GTGGAGGNGG PGGDGGTGAT VGFAGSGGFG 420
GAGGIAQLFG TGGMGGSGGG IGAGTTTVVP PDVAPVGGTG GNGGRAGLLL GVGMGGNGG 480
ATSVGGTLYA AGNGGDGGL VWGNGGTGGS GGAGGAGSVG NGGAGGNAAL LFGNGGAGGA 540
GGAGGIGAGG AGGFGAVLFG NGGAGGSGAP GGIGAGGNGG NALLVNGGN GGAGTGAAG 600
30 GAGGSGGLLF GQNGMPGP 618
<212> Type : PRT
<211> Length : 618
SequenceName : SEQ ID 150
SequenceDescription :
35 Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
40 MNFSLPPEI NSALIFAGAG PEPMAAATA WDGLAMELAS AAASFSGSVTS GLVGGAWQGA 60
SSSMAAATA PYAAWLAAA VQAEQTAAQA AAMIAEFEAV KTAVVQPM LV AANRADLVSL 120
VMSNLFGQNA PAIAAIEATY EQMWAADVSA MSAYHAGASA IASALSPFSK PLQNLAGLPA 180
WLASGAPAAA MTAAAGIPAL AGGPTAINLG IANVGGGNGV NANNGLANIG NANLGNYNFG 240
SGNFCSNIG SASLGNNGN FGNLGSNNVG VGNLGNLNTG FANTGLGNFG FGNTGNMNG 300
45 IGLTGNNQIG IGLLNSGTGN FGLFNSGSGN VGFFNSGNGN FGIGNSGNEN TGGWNSGHGN 360
TGFFNAGSFN TGMLDVGNAN TGSNTGSGYN MGDFNPGSSN TGTFNTGNAN TGFLNAGNIN 420
TGVFNIGHMN NGLFNTGDMN NGVFYRGVQ GSLQFSITTP DLTLPPLQIP GISVPAFSLP 480
AITLPSLNIP AATTPANITV GAFSLPGLTL PSLNIPAAAT PANITVGAFS LPGLTLPSLN 540
IPAATTPANI TVGAFSLPGL TLPSLNIPAA TTPANITVGA FSLPGLTLPS LNIPAAATPA 600
50 NITVGAFSLP GLTLPSLNIP AATTPANITV SGFQLPPLSI PSVAIPPVTV PPITVGAFNL 660
PPLQIPEVTI PQLTIPAGIT IGGFSLPAIH TQPIITVGQIG VQFGLPSIG WDVFLSTPRI 720
TVPAFGIPFT LQFQTNVPAL QPPGGGLSTF TNGALIFGEF DLPQLVHPY TLTGPIVIGS 780
FFLPAFNIPG IDVPAINVDG FTLPQITTPA ITTPEFAIPP IGVGFTLPQ ITTQEIITPE 840
LTINSIGVGG FTLPQITTP ITTPTLTIDP INLTGFTLPQ ITTPTITTP LTIDPINLTG 900
55 FTLPQITTP ITTPTLTIEP IGVGFTTP LTVPGIHLPS TTIGAFAPG GPGYFNSSTA 960
PSSGFFNSGA GGN SGFGNNG SGLSGWFNTN PAGLLGGSGY QNFGGLSSGF SNLGSVSGF 1020
ANRGILPFSV ASVVSGFANI GTNLGFFQG TTS 1053
<212> Type : PRT
<211> Length : 1053
SequenceName : SEQ ID 151
SequenceDescription :
60 Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
65 MLYVVASPD LMTAAATNLAE IGSAISTANG AAALPTVEVV AAADEVSTQ IAAALFGAHR 60

SYQTLSTQAA AFHSRFVQAL TTAAASYASV EAANASPLQV ALDVINAPAQ TLLGRPLIGN 120
GADGSTPGQA GPGGGLLYGN GGNGAAGGPN QAGGAGGNAG LIGNGGAGGA GGVGAVGGKR 180
GTGGLLFGNG GAGGQGGGLG AGINGGSGGQ GGHGGNAILF GQGGAGGPGG TGAMGVAGTN 240
PTPIGTAAFG SDGVNQIGNG GNTDLTGGAG GDGNAGSTTV NGNGGTTGGA ARNSSGGTGN 300
5 SFGGAGGAGG DGANGGDGGA GGEALTEGGA TAVSGAGGKG GNAEASGGAG GNGGKGGFAQ 360
ATTSVTGGNG GNGGNHDSN APGGAGGSGG VGGDGGRGGL LAGNGGTGGA GGNGGTGGAG 420
APGGAGGAGG KADIANSIGD NATVTGGNGG TGGDGGGALG TGGAGGAGGL GGHGAGGGLL 480
IGNGGAGGAG GLGGAGGAGG AGGEGGAGGA GGEAIPGGAS TNSAGGDGGA GGTGCGNGGDG 540
GAGGAPGLGG AGGAGGWLIG QSGSTGGGGA GGAGGAGGAG GAGGSGGAGG HGDITSKNG 600
10 SSGTAGFDGN PGQPG 615
<212> Type : PRT
<211> Length : 615
SequenceName : SEQ ID 152
SequenceDescription :
15
Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
20 MHYSVLPEI NSALIFAGAG SGPMLAAASA WDGLATELAS AAVSFGSVTA GLVGGSWQGR 60
SSVMAAAAA PYAGWLAAAA TQAEQAATQA QVMVAEFEAV RLAMVQPALV AANRSGLISL 120
VISNLFQNA PAIAAAEAAY EEMWALDVSA MAAYHSGASA VAVALPAFAL PLRLPAGLAA 180
GPAAVVTALT TAVGMPFAG RAIAASLGLA NVGGGNLGNNA NNGLGNIGNA NLGNNNLGSG 240
NFGSFNIGSA NLGGNNIGIG NAGANNFGLA NLGNLNTGFA NAGIGNFGIA NTGNNNIGNG 300
25 LTGNNQIGIG GLNSGNGNVG LFNAGSANIG FPNSGNGNFG IGNSGNFSTG LFNPGHGNTG 360
FLNAGSFNTG MFDVGNANTG SFNVGHYNGF AFNPGPSNTG TFNTGGANTG WFNTGSINTG 420
AFNIGDMNNG LFNTGDMNNG VFYRGVQGS LQFAITSPDL TLPSLEIPGI SVPAFSLPAI 480
TLPSLTIPAV TTPANVTGA FDLPLGLTVPS LTPAAMTPA NITVGAFDLP GLTVPSLTIP 540
ATTPANITV GAFNLPQLSI PSVTVPPI TI PAGTALGAFN LPTLSIPSVT VPPITIPAGT 600
30 TVGGFTLPTI HTPLISTPQI SIGGFSTPGI ATQANSQVIN LPTFSLNGIT ITNLVVFIPN 660
NITALQTNMP GVFPQIGGFA NTPPAFINTG TITVGGGQIN GVGFSIGAIN VTPFTLENVV 720
IQPWSLGGIS VDGFTLPEIS TQEFITPAIT ISPIGVGALS LPDITTTQFT TPELTIDPIT 780
LGGFTLPQLS IPAITTPAFT IDPIALGGFT TPQIMTPEIT TPPFAIDPIG LSGFTLPQVN 840
IPBITTPEFT IQPVGLAAFT TPALTIASIH LPSTTMGGFA IPAGPGYFNS SATPSLGFFN 900
35 AGIGCNSGFG NSGSGLSGWF NTSPVGLLAG SGYQNYGGLI SGFSNLGSGI SGFANTGTLP 960
FAVTSLVSGL ANIGNLNLGL FQSTTP 987
<212> Type : PRT
<211> Length : 987
SequenceName : SEQ ID 153
SequenceDescription :
40
Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
45 MSFVVVAPEV LAAAASDLAG IGSTLAQANA AALAPTTAVL AAGADEVSAA IASLFGAHGQ 60
AYQAVSAQMS AFHAQFMQAL TGAGGAYAAA EAVNVSAQAS VEQDLLAATN ARFERIFGRP 120
LIGDGANGGP QDGGPGGGLL YGNGGNGGTS TTVGMAGNGG GAAGLIGNGG FGGGGGPGAA 180
GGNGGAGGWL FGNGGAGGAG GLGVAPGVPG GAGGAGGAGG VGGPAGLWGH GGAGGAGGAG 240
50 VAGAGGFEGT IGAGGAGGVG GAGGVGGAGG AGGWLYGDAG AGGDGGVGGG GGTGGLGNRG 300
GAGGAGGAGG VGGAGGAAGL WGGGGAGGVG GTGGGAGLGA QSVTFSSSLG GLSGGDGGAG 360
GAGGAGGAGG TGGWLYGGGG AAGSGGDDGT GQGGGAGGAG VFSLFGSGGG PGGNGGVGGV 420
GGVGGAGGRA GLFGVGLGG AGGDAGDSGE GGFGGPGLAG GLFGNPGNGG VGGIGGDAAG 480
GGAGGAGGNG GAGGNGWLFG GNGGAGGSGG DGGAGAGGGA GNLGSAGGIN APAGNPGSGS 540
55 VGIGGAGGAG GTAGLFGDGG AGGAGGAGAA GFGGGISAAT PSAGSEGAMG GAGGVGNAR 600
LLGTGGAGGV GGGGGAGGDG GRGGVATPGG QGGDAGDGA GAGGNGGGA SGAGGWLLGT 660
GGAGGAGGNG GNGKAGFSP GPTNFGLNGA GGGGVGGNG ATGPWLFGDG GPTPGSTGAG 720
AAGGHGGDAQ LIGNGGHGA GGTGVPNGSG GAGGLSGLLF GEPGANG 767
<212> Type : PRT
60 <211> Length : 767
SequenceName : SEQ ID 154
SequenceDescription :
65
Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :

MSFVIANPEM LAAAAATDLAG IRSAlSAATA AAAAPTlQVA AAGADEVSLA ISALFGQHAQ 60
AYQALSAQAT IFHDQFVQAL TSGGNLYAAA ESHTVEQMVL NAINAPTQTL FGRPLIGDGA 120
NGTAENPDGQ NGGLLFGNGG NGFTQTTAGV AGGNGGSAGL IGNGGAGGGG GAGAAGGLGG 180
5 NGGWLYGNGG AGGIGGAGTG TGGHGGAGGA GGRRAWLWGTG GAGGAGGDGG WLFQDGGAGG 240
TGGNGGSGFN SLTSSVGCAG GAGGHAGLFG AGGTGGTGGI GGQNTETGPA ASNGGAGGAG 300
GGGGYLVGDG GAGGTGGAGG KNSSGGATLT GGTGGTGGAG GAAGWLYGSG GAGGAGGAGG 360
LNNAGGATGG TGGTGGAGGS GAWLYGNGGA AGAGGNGGNN TSAGTGGVGA SGCTGGNAGL 420
IGAGGHGGAG GAGGNQTGGV GNGGAGGNGG AGGAGGQLYG NGGDGGNGGA GGANIAGGNG 480
SDGGAAGHGG AGGSARLIGA GGHGGDGGAG GNTAGRRADA IAGTGGDGGN GGNGLLSGN 540
10 AGAGGHGGAG GSSTATTTTG TPPTGATGGN GGNCGAGGTA GFTGSGGIGG NGGAGGTGGN 600
AGVALSVGST GGLGGNGGSG GLGGGGGSLF GNGGAGGVGA TGGNGGSGIG PASVGGNGGK 660
GGVGAAGGLA GQIGNGGSGG SSGAGGNGGT GDTAGNGNGG GAGAVGGNAQ LIGNGGNGGG 720
CGNGGTGADG T 731
<212> Type : PRT
15 <211> Length : 731
SequenceName : SEQ ID 155
SequenceDescription :

Sequence

20 <213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MPGRFRNFGS QNLGSGNIGS TNVGSNGNIGS TNVGSNGNIGD TNFGNGNNGN FNFGSGNTGS 60
NNIGFGNTGS GNFGFGNTGN NNIGIGLTGD GQIGIGGLNS GSGNIGFGNS GTGNVGLFNS 120
25 GTGNVGFNGS GTANTGFGNA GNVNTGFWNG GSTNTGLANA GAGNTGFFDA GNYNFGSLNA 180
GNINSSFGNS GDGNSGFLNA GDVNSGVNA GDVNTGLGNS GNINTGGFNP GTLNTGFFSA 240
MTQAGPNSGF FNAGTNSGF GHNDPAGSGN SGIQNSGFNG SGYVNTSTTS MFGGNSGVLN 300
TGYGNSGFYN AAVNNTGIFV TGVMSGFFN FTGNSGLLV SGNGLSGFFK NLFG 354

30 <212> Type : PRT
<211> Length : 354
SequenceName : SEQ ID 156
SequenceDescription :

Sequence

35 <213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSFVLAMPEV LGSAAATDLAA LGSVLGAADA AAAATTTGIV AAAQDEVSA IAALFSAHGR 60
40 AYQVASAQAA AVHAQFVEAL SAGAGAYASA EAAGAAVLAN PAQSVQQDIL AAVNAQSVAL 120
TGRPLIGNGA NGAPGTGANG APGGWLLGNG GAGGSAAAGS GLPGGAGGAA GLFGTGGAGG 180
AGGSSTVGDG EAGGAGGSGG WLLGTGGVGG VGLGAGAGG AGGVGGAGGL LGAGGHGGAG 240
GLGAVTGGVG GTGGAGLLA GLLAGPGGAG GTGGRGFLNN GGVGGAGGNA GLLFGAGGTG 300
GSGGAGLGGD GGAGGAGGNT GVLFGNAGSG GTGGFGDTDG GAGGAGGDAG WLGSAGVGG 360
45 GGFGETDGG VGGAGGKAGL LIGNGGAGGA GGQAVTGGT GGAGGDGVL I GNGGNAGIGG 420
TGPTAGDTGA GGISGLLLGA DGFNTPASAS PLHTLKQQA AAINAPTQTL TGRPLIGNGT 480
PGAVSGCATG APGGWLLGDC GAGGSGAAGS GAPGGAGGAA GLWGTGGAGG AGGSSAGGGG 540
AGGAGGAGGW LLGDGGAGGI GGASTVLGGT GGGGGVGGW GAGGAGGAGG TGLVGGDGA 600
GGAGGTGGLL AGLIGAGGGH GGTGGLSTNG DGGVGGAGGN AGMLAGPGA GGAGGDGENL 660
50 DTGGDGGAGG SAGLLFGSGG AGGAGGFGFL GGDGAGGNA GLLSSGGAG GFGGFGTAGG 720
VGGAGGNAGW LGFGGAGGVG GSAGLIGTGG NGGNGGTGAN AGSPGTGGAG GLLLGQNGLN 780
GLP 783
<212> Type : PRT
<211> Length : 783
55 SequenceName : SEQ ID 157
SequenceDescription :

Sequence

60 <213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSLVIATPQL LATAALDLAS IGSQVSAANA AAAMPTTEVV AAAADEVSAA IAGLFGAHAR 60
QYQALSVQVA AFHEQFVQAL TAAAGRYAST EAAVERSLLG AVNAPTEALL GRPLIGNGAD 120
GTAPGQPGA GGLLFGNGGN GAAGGFGQTG GSGGAAGLIG NGGNGGAGGT GAAGGAGGNG 180
65 GWLWNGCGNG GVGGSVAAG IGGAGGNGGN AGLFGHGGAG GTGGAGLAGA NGVNPTPGA 240
ASTGSPADV SGIGDQTGGD GGTGGHGTAG TPTGTTGGD ATATAGSGKA TGGAGGDGTT 300
AAAGGGGNG GDGGVAQDI ASAFGGDGGN GSDGVAAGSG GSGGAGGGA FVHIATATST 360

GGSGGFGGNG AASAASGADG GAGGAGGNGG AGGLLFGDGG NGGAGGAGGI GGDGATGGPG 420
GSGGNAGIAR FDSPDPEAEP DVVGGKGGDG GKGGSGLGVG GAGGTGGAGG NGGAGGLLFG 480
NGGNGGNAGA GGDGGAGVAG GVGNGGGGGG TATFHEDPVA GVWAVGGVGG DGGSGGSSLG 540
VGGVGGAGGV GKGKGASGML IGNGGNGGSG GVGAGGAGVG AGGDGGNGGS GGNASTFGDE 600
5 NSIGGAGGTG GNGGNGANGG NGGAGGIAGG AGGSGGFLSG AAGVSGADGI GGAGGAGGAG 660
GAGGSGGEAG AGGLTNGPGS PGVSGTEGMA GAPG 694
<212> Type : PRT
<211> Length : 694
SequenceName : SEQ ID 158
10 SequenceDescription :

Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
15 <400> PreSequenceString :
MSFVIAAPEV IAAAATDLAS LESSIAAANA AAAANTTALL AAGADEVSTA VAALFGAHGQ 60
AYQALSAQAQ AFHAQFVQAL TSGGGAYAAA EAAATSPLLA PINEFFLANT GRPLIGNGTN 120
GAPGTGANGG DGGWLIIGNG AGGSGAAGVN GGAGNGGAG GLIGNGGAGG AGGRASTGTG 180
GAGGAGGAAG MLFGAAGVGG PGGFAAFAA TGGAGGAGGN GGLFADGGVG GAGGATDAGT 240
20 GGAGGSGENG GLFGAGGTGG PGGFIFGGG AGGDGSGSGL FGAGGTGSGS GTSIINVGGN 300
GGAGGDAGML SLGAAGGAGG SGGSNPDGGG GAGGIGDGG TLFGSGGAGG VCGLGFDAGG 360
AGGAGGKAGL LIGAGGAGGA GGGSPAGAGG TGGAGGAPGL VGNAGNGGNG GASANGAGAA 420
GGAGGSGVLI GNGGNGSGSG TGAPAGTAGA GGLGGQLLGR DGFNAPASTP LHTLQQIILN 480
AINEPTQALT GRPLIGNGAN GTPGTGADGG AGGWLFNGG NGGHGATGAD GGDGSGSGAG 540
25 GILSGIGTGT GSGGIGTTGQ GGTGGTGGAA LLIGSGGTGG SGGFGLDTGG AGGRGGDAGL 600
FLGAAGTGGQ AALSQNFIGA GGTAGAGGTG GLFANGGAGG AGGFGANGGT GGNGLLFGAG 660
GTGGAGTLGA DGGAGGHGGL FGAGGTGGAG GSSGCTFGGN GSGGNAGLL ALGASGGAGG 720
SGGSALNVGG TGGVGGNGGS GGSLLFGFGA GGTGSSGIG SSGGTGGDGG TAGVFGNGGD 780
GGAGGFGADT GGNSSVPNA VLIGNGNGG NGKAGGTPG AGGTSGLIIG ENGLNGL 837
30
<212> Type : PRT
<211> Length : 837
SequenceName : SEQ ID 159
SequenceDescription :
35

Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
40 <400> PreSequenceString :
MSFVIAVPET IAAAATDLAD LGSTIAGANA AAAANTTSLL AAGADEISAA IAALFGAHGR 60
AYQAASAEAA AFHGRFVQAL TTGGGAYAAA EAAAVTPLLN SINAPVLAAT GRPLIGNGAN 120
GAPGTGANGG DAGWLIIGNG AGGSGAKGAN GGAGGPGGAA GLFGNGGAGG AGGTATANNG 180
IGGAGGAGGS AMLFGAGGAG GAGGAATSLV GGTGGTGGTG GNAGMLAGAA GAGGAGGFSF 240
STAGGAGGAG GAGGLFTTGG VGGAGGQCHT GGAGGAGGAG GLFGAGGMGG AGGFGDHGTL 300
45 GTGGAGGDGG GGLFGAGGD GGAGGSLTT GGAAGNGGNA GTLSLGAAGG AGGTGGAGGT 360
VFGGKGGGAG GAGGNAGMLF GSGGGGGTGG FGFAAGGQGG VGSAGMLSG SGGSGGAGGS 420
GGPAGTAAGG AGGAGGAPGL IGNGGNGGNG GESGCTGGVG GAGGNAVLTG NGGEGGIGAL 480
AGKSGFGGFG GLLLGADGYN APESTSPWHN LQDDILSFIN EPTEALTGRP LIGNGDSGTP 540
GTGDDGAGG WLFNGGNGG AGAAGTNGSA GGAGGAGGIL FGTGGAGGAG GVGTAAGGA 600
50 GGAGGSAFLI GSGGTGGVGG AATTGGVGG AGGNAGLLIG AAGLGGCGGG AFTAGVTTGG 660
AGGTGGAAGL FANGGAGGAG GTGSTAGGAG GAGGAGGLYA HGGTGGPGGN GGSTGAGGTG 720
GAGGPGGLYG AGGSGGAGGH GGMAGGGGGV GGNAGSLTLN ASGGAGGSGG SLSGKAGAG 780
GAGGASAGLFY GSGGAGGNGG YSLNGTGGDG GTGGAGQITG LRSFGGAGG AGGASDTGAG 840
GNGGAGGKAG LYNGGDDGGA GGDGATSGKG GAGGNAVVG NGGNGGNAGK AGGTAGAGGA 900
55 GGLVLGRDQ HGLT 914
<212> Type : PRT
<211> Length : 914
SequenceName : SEQ ID 160
SequenceDescription :
60

Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
65 <400> PreSequenceString :
MSLVIVTPET VAAAASDVAR IGSSIGVANS AAAGSTTSVL AAGADEVSAA IATLFGSHAR 60
EYQAISTQVA AFHDRFAQTL SAAVGSYVSA EATNAAPLAT LEHNVNLN LN APTQALLGRP 120
LIGDGAAGAP GTGQAGGAGG ILWNGGAGG SGAPGQVGA GGAAGLFGTG GAGGAGGAGA 180

AGGAGGSSGGW LLGNGGVGGA GGQSLGGAT GGAGGNAGLF GVGGTGGPGG PGGPGGVGGT 240
GGAGGLGGTL YGAGGHGGAG GPGPIGGVGG HGGVGGAGL LGVGGHGGAG GHGAEGVAGA 300
AGEDLSPHGT SGGVGGDAGD GGTGGRGGWL AGAGGAGCAG GVGGTGGAGG AGFSRALIVA 360
5 GDNNGDGGNG GMGGAGGAGG PGGAGGLISL LGGQGAGGAG GTGGAGGVGG DRGAGGPGNQ 420
AFNAGAGGAG GHGGDPGAGG AGGTGGAGSI TGAQGAIGAT PTSGGNGCAG GNGANATTAG 480
TNGANGGPGG HGGLVGNCGA GGNANGAAG TNASDSCAVG GKGNSSGNGG QGGAGGDGGT 540
LAGNGGAGGT GGRGADGGLG GSGAEGANAT TAGERGQDGG KGGNGGVGGT GGNVAVPGAN 600
GGHGGNGGNP GFSGAGGLGG LSGDGVTRAA QGATPDFADT GKGKNGGNG ANAVAPGGTG 660
ASGGAGGNAG AGGKGGENII GDGGGGNGGA GKGKGAGTLL GLTVFGDNNG AGVLGDSTDP 720
10 DSGGAGGAG GAGGAGDPT I 741
<212> Type : PRT
<211> Length : 741
SequenceName : SEQ ID 161
SequenceDescription :
15
Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
20 MSFVTAAPEM LATAAQNVAN IGTSLSAANA TAAASTTSVL AAGADEVSQA IARLFSYAT 60
HYQSLNAQAA AFHHSFVQTL NAAGGAYSSA EAANASAQAL EQNLLAVINA PAQALFGRPL 120
IGNANGTAA SPNGGDGGIL YNGGNGFSQ TTAGVAGGAG GSAGLIGNGG NGGAGGAGAA 180
GGAGGAGGWL LGNGGAGGPG GPTDVPAGTG GAGGAGGDAP LIGWGGNGGP GGFAPFGNGG 240
AGNGGASGS LFGVGGAGGV GGSSEVDVGT GAGGAGGRGL FLGLGGDGA GGTNNNGGD 300
25 GGAGGTAGGR LFSLGGDGN GAGTAIGSN AGDGGAGGS SALIGYAQGG SGGLGGFES 360
TGDDGGLGGA GAVLIGTVG GFGGLGGSN GTGGAGGAG TGATLIGLGA GGGGGIGGFA 420
VNVNGVGGGL GQCGQGAAL IGLGAGGAG AGCATVVLG GNGGDGGDG GLFSIGVGGD 480
GGNANGAMP ANGGNGGNAG VIANGSFAPS FVGFGNGGN GVNGGTGSG GILFGANGAN 540
GPS 543
30 <212> Type : PRT
<211> Length : 543
SequenceName : SEQ ID 162
SequenceDescription :
35
Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
40 MSYVLATPEM VAAANNLAQ IGSTLSAANA AALAPTGTVL AAGADEVSAA VASLFSGHAQ 60
AYQTLTQAA AFHERFIQAL STAAGAYGSA EAANASPLQQ ALNVINAPTQ TLLGRPLIGN 120
GTNGAPGTGQ AGGPGGLLYG NGGNGSGGV GQAGGAGSA GLIGIGGTGG AGGAGAVGGV 180
GGNGGWLYGN GGAGGLGGTG VAGVNGMGMA AGGAGGNAYL FGSGGAGGQG GMGAAGADGV 240
NPTPTGTADA GSTGTDQTLG GNAIGNGGP GDAGDAMTSG GAGSGGNAV STVNGDAVGG 300
EGKKGEGAY GGAGGAGGSA ASIGNAAIG NGGAGGNAQA PGGVGGAGGE GGDQVGTNS 360
45 PSNAEAGNGG SGGNGFDSFA SGTGGAGGT GGAGGRGGL IGDGGAGGAG GVGTTGSSGA 420
PGGGGAGGD GGAANTDSAG SSRKAFGGDG GVGDDGASAL GTGEGGGIGG QGGNGGAGGL 480
LIGNGGAGGV GGTAGAGGTG GSGGAGGAG AGGGGTNSGP GAAPGNGNT GNGGNGGAP 540
GALGGKGGSG GLIGRAGSDG GVAGGAGGA GGAGGTGGE GTGGDGKTTD GNPMMGGSPG 600
SAGQPG 606
50 <212> Type : PRT
<211> Length : 606
SequenceName : SEQ ID 163
SequenceDescription :
55
Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
60 MSFLFAQPEM LGAAATDLAS IGSAISTANA AAAAATTRVL AAGADEVSAA VAALFSGHAQ 60
TYQALRTQAA AFHQIIVQTL TSTAGAYASA EAANVEQQLL GAINAPTMA LGRPLIGHGA 120
DGAPGTGQAG GAGGILYNG GNGSGATGQ AGGAGGAAGL IGHGGAGGLG GTGASGGAGG 180
AGGWLWNGG AGNGGVGVA GPPGVGGAG GAGGAAGLWG SGGSGGTGCG GGVGGKSGD 240
GGTGGIGGAG GGGWLHGDG GAGGHGGQG TGVSSGNGG AGGTGGDGRG LSGSGAGGR 300
GGQTGVGKV GENNFGGAGG AGGTGGLIGN GGAGNGGQG AISGAGGAG NAWLIGDGA 360
65 GNGGDIRGQ GGGAGGAGGA GGQLIGNGT GGAGGTVTSP NGLGGAGGAG GSAGLIGHG 420
TGGAGGHSAG GPDGNGGI GG AGGAGNGGQ LYGTGGTGGT GKGGDGFGV FGKGGAGGTG 480
GRGGAAGLIG DAGTGGTGGK GGTAGEDGTG GNGGTGNGG AAVLIGNGGG GGAGNGGAG 540

NDGTPGNGGG GGVGGTGGTL FGQPGQPGPP GQPGPA 576
<212> Type : PRT
<211> Length : 576
SequenceName : SEQ ID 164
5 SequenceDescription :

Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
10 <400> PreSequenceString :
MWTSQMIVAP AFVDAAAKDL ATIGSAISRA NAEALVPITA LLPAGADDVS AAIAALFATH 60
QQAQELSAH AVAFHEQFVQ LMSAGAAQYA SAEANSSPL QIVGQTALDA INSPVQTLTG 120
RPLIGNGANG VAGTGQNGGD GGLYGNNGN GSGGTGQNG GNGGSAGLWG SGGNGGQGG 180
GANGAAGQPG KAGGSGGNGG AGGWIYGHGG HGGAGNGGCM ATAPCGASAG FDGGACNNGG 240
15 SGGRGGLLFG NGGNGSVGGM GGQGTNDTAG DSAGSGGLGG NGGNGAQGGW LIGNGGQGGD 300
SGAGGTDST QTGVMMNGASG GSAGIAGNGG DAGLVNGGA GNGGNGAAG SALGTTIFGG 360
SGGVGSGGGD GGNGGWLFGS GASGNGGQG GDAGTNGFAG FGGSAGGGGW VGAVNFGPIS 420
VQGFGLFHHG GDGGNGGDDV AGSLSIQFGA SGGDGGQGGV LYNGGNGGN AGSGGGTGFE 480
GSAGQGGAAI LIGNGGAGGN GATGTTGVGN IIQEAGDGS DGGAGGSGGL LFGSGGAGGI 540
20 GGAGGVGSG NDGGNGGDDG QGGASGLGIG NGGPGSGGT GGAGGTGSSA GTGGAGGDDG 600
NAALLIGTGG DGGDGVPPAP GGQGGKGLI GLPGQNGQP 639
<212> Type : PRT
<211> Length : 639
SequenceName : SEQ ID 165
25 SequenceDescription :

Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
30 <400> PreSequenceString :
MSWVMVSPLE VVAAAADLAG IGSASSANA AAANTTTGLL TAGADEVSTA IAALFGAQQG 60
AYQAASAQAA AFYAQFVQAL SAGGGAYAAA EAAAVSPLA PINAQFVAAT GRPLIGNGAN 120
GAPGTGANGG PGGWLIGNGG AGSGGAPGAG AGGNGGAGGL FGSGGAGGAS TDVAGGAGGA 180
GGAGGNAGML FGAAGVGGVG GFSNGGATGC AGGAGGAGGL FGAGRERGS GSGNLTGGAG 240
35 GAGGNAGTLA TGDGGAGGTG GASRSGGFAG AGGAGGDAGM FFGSGSGGGA GGISKSVGDS 300
AAGGAGGAPG LIGNGGNGCN GCASTGGGDG GPGGAGGTGV LIGNGGNGGS GGTGATLGKA 360
GIGGTGGVLL GLDGFTAPAS TSPLHTLQD VINMVNDPFQ TLTGRPLIGN GANTPTGTGA 420
DGGAGGWLFG NGGNGGQGTI GGVNGGAGGA GGAGGILFGT GGTGSGGGPG ATGLGGIGGA 480
GGAALLFGSG GAGGSGGAGA VGGNGGAGGN AGALLGAAGA GGAGGAGAVG GNGGAGGNGG 540
40 LFANGGAGGP GGFSGPAGAG GIGGAGGNGG LFGAGGTGGA GGGSTLAGGA GGAGNGGLF 600
GAGGTGAGS HSTAAGVSGG AGGAGGDAGL LSLGASGGAG GSGGSSLTAA GVVGGTGGAG 660
GLLFGSGGAG GSGGFSNSGN GGAGGAGGDA GLLVSGGAG GAGASATGAA TGGDGGAGGK 720
SGAFGLGGD GAGGATGLSG AFHIGGKGV GGSAVLIGNG GNGGNGNSG NAGKSGGAPG 780
PSGAGGAGGL LLGENGLNGL M 801
45 <212> Type : PRT
<211> Length : 801
SequenceName : SEQ ID 166
SequenceDescription :

50 Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
GQSYQAVSAQ AAFAHDFVQ LLNAGGGSYA SAEIANAQQN LLNAVNAPTQ TLLGRPLVGD 60
55 GADGASGPVG QPGGDDGILW GNGGNGDST SPGVAGGAGG SAGLIGNGGR GGNGAPGGAG 120
GNGGLGGLL GNGGAGGVGG TGDNGVDDL GGGGGDGG GGRAGLIGHG GAGGNGGDGG 180
HGGSGKAGGS GSGGFGQFG GAGGLLYNG GAAGSGNGG DAGTGVSSDG FAGLGGSGGR 240
GGDAGLIGVG GGGGNGGDP GLGARLFQVG SRGGDGGVGG WLYDGGGGG DGGNGGLPFI 300
GSTNAGNGS ARLIGNGGAG GSGSGAPGS VSSGGVGGAG NPGSGGNGG VWYNGGAGG 360
60 AAGQGGPGMN TSPGGPGGV GGHGGTAILF GDGGAGGAGA AGGPPTPDGA AGPGSGSGTG 420
GLLFGVPGPS GPDG 434
<212> Type : PRT
<211> Length : 434
SequenceName : SEQ ID 167
65 SequenceDescription :

Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
5 MAHFSVLPPE INSLRMYLGA GSAPMLQAAA AWDGLAAELG TAASSFSSVT TGLTGQAWQG 60
PASAAMAAAA APYAGFLT TA SAQAQLAAGQ AKAVASVFEA AKAAIVPPAA VAANREAF LA 120
LIRSNWLG LN APWIAAVESL YEEYWAADVA AMTGYHAGAS QAAAQLPLPA GLQQFLNTLP 180
NLGIGNQGNA NLGGGNTGSG NIGNGNKGSS NLGGGNIGNN NIGSGNRGSD NFGAGNVGTG 240
NIGFGNQGP I DVNLLATPGQ NNVGLGNIGN NNMFGFNTGD ANTGGGNTGN GNIGGGNTGN 300
NNFGFNTGN NNIGIGLTGN NQMGINLAGL LNSGSGNIGI GNSGTNNIGL FNSGSGNIGV 360
10 FNTGANTLVP GDLNNLG VGN SGNANIGFGN AGVLNTGFGN ASILNTGLGN AGELNTGFGN 420
AGFVNTGFDN SGNVNTGN GN SGNINTGSWN AGNVNTGFGI ITDSGLTNSG FGNTGTDVSG 480
FFNTPTGPLA VDVSGFFNTA SGGTVINGQT SGIGNIGVPG TLFSGSVRSGL NTGLFNMGT A 540
ISGLFNLRLQ LG 552
<212> Type : PRT
15 <211> Length : 552
SequenceName : SEQ ID 168
SequenceDescription :

Sequence
20 -----
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
25 MSFLIASPEA LAATATYLTG IGSATSAANA VAAAPTTEIL AAGTDEVSTA ISALFGAHAQ 60
AYQALSAHVA AFHDQFVHTL TAGAGSYMAA EAAAASPLQA LQLELLNAIN APTLALLGRP 120
LIGDGTDAAP GSGGAGGAGG ILIGNGGTGG ASDLAGTGRG GVGAGGAGAG LFGIGGAGGG 180
CGSAVAIGGD GGAGGAGGVF SGGGAGGAGD AIGGSGGAGG TGGLLG GGGG AGGAGGAGGN 240
GGGASNSASI GGDGGSGGAG GMLYGAGGVG GNGGAAVAIG GDGGAGGRAG AIGNGGDGGN 300
GGTSNTPGGS GGDGGNGGNA GLIGNGGNGG NAEIVISGGS VAGTGGNGGL LLGPNGTNGL 360
P 361
30 <212> Type : PRT
<211> Length : 361
SequenceName : SEQ ID 169
SequenceDescription :

Sequence
35 -----
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
40 AQASPAAHGG SGGAGGNGGA GSAGNGGAGG AGGNGGAGGN GGGGDAGNAG SGGNGGKGGD 60
GVGPGSTGGA GKGKGAGANG GSSNGNARGG NAGNGGHGGA GSGGDTGGAG GAGGQGGFGG 120
TGGSGSGIGG GAGGNGGNGG AGGTGVVLGG KGGDGGNGDH GGPATNPGSG SRGGAGGSGG 180
NGGAGGNATG SGGKGGAGGN GGDGSFGATS GPASIGVTGA PGGNGGKGGGA GGSNPNNGSGG 240
DGGKGGNGGA GGNNGSIGAN SGIVGGSGGA GGAGGAGNG SLSSGEGGKG GDGGHGGDGV 300
GGNSSVTQGG SGGGGGAGGA GSGGFFGGKG GFGGDGGQGG PNGGGTVGTV AGGGGNGGVG 360
45 GRGGDGVFAG AGGQGGGLGG GNGGGGSGTG NGGLGGAGG GGNAPDGGFG GNGGKGGQGG 420
IGGGTQSATG LGGDGGDGGD GGNGGNSGAK AGGAGGKGQA GQPNSGTEPG FGGDGGGLGA 480
GATP 484
<212> Type : PRT
<211> Length : 484
50 SequenceName : SEQ ID 170
SequenceDescription :

Sequence
55 -----
<213> OrganismName : Rickettsia prowazekii
<400> PreSequenceString :
MKKSKILRKF LATASLCGTL FTNSNATGTI IPNNGSVSLN TDAGLVGGVF NNGDIIQIVN 60
GGREIKISAD KANAIIGGIN TLKELPDFGG VEVSONVSIG PLNAGEDLNT NFGPLKFISN 120
NVTSIITGVG TKTF SNIDFA GKNA TLQINK DLNITTKIDN TVAGNNGSIT FEGSGIISNH 180
60 IGYTNSLLGI NVNGEAKIY APEANNITIN AKNINLTHNN SILTLC DGN I TTLKGNINNT 240
TEIDGQGILN LAYDLGSSSI ITGDIGNIGS LDTINVLGS ATFNSTILKA TNINLKHNTS 300
TLNLDNIIIV IGNIKGNNNK DILNFKVHGT NLDNEMIIPA POKTHGT LNF KGNATLNGNI 360
NNLNILKFSG GHGKTLNLQG NTKVDNLVFA DSVLDSGTIS VNGLLD TDCV TFNNSNVNGG 420
TLIINAKNTI SAKLLNATKA KIQINANLTM NHPSAGDISD IRIADNTIYT IDAKNGNVNL 480
65 LNNNAKIIFE GADSM LALIN TGV TADRTFT IYNNLNQSGN DEYGIVKIEA IKKVITIANQ 540
SGPYTIQDN THRLKELIVE GAGDIIIDDT IFTKLLSINS TGQITFNRTL DLGAGGNIAF 600
GKHGTLVVNG VTGSITTSEN NQGILTINS NITGVIGTNE LGLKLVNIGA DPVTCSANVF 660

	ASVALTNPSS	VLILADGVTL	TGEVTTNNNT	KGVLSLGTGS	NITQIGTNS	AALEKINIGA	720
	GASNIDSNIY	AGSTVLTDQT	SELTNNNDVV	VNSNIITTAG	NNSGKLITG	NGGITGNIGA	780
	NGAALQEVVF	NGTTNIGGTA	NSQNFVAHS	AANVVITGLT	TGALKYKDTG	TIIAHGGLVG	840
	DIDFNNKAGK	FILGDGAMID	GSVLCNGGVA	GTLDFIGDGN	VTQNIAGDNA	NSISTINIQQ	900
5	DNTKNVTIAN	DIFVDNIHFT	NGGILQLGCG	LTTHNIDFGA	NGGTLEFNGN	NTYNLNAIIV	960
	NGQNGILNAF	TNLKASDDTI	GTVKIINIGQ	IGTPQNFTIQ	VNNKNLTLVS	SVNSSINFGD	1020
	ANSQILSAP	VDQTIKFINN	LNETGGGIIT	LDSNGNNLTI	SGNNGIKLGS	KGNELSSLNI	1080
	KGKVTVTNDL	DIQNIHQNLN	NNGALFDDQS	LTSAKIKNIN	IGTVAGGATY	TLDAINDNFD	1140
	LNTSGMVFKH	QDSILELKNS	SNTNDHTITL	TSALDPGNNQ	FGIIKLITDT	NKLTIDNNGN	1200
10	VAYTLGTANH	MLKQLTFASI	DNGAIALKVG	INVENVTLNI	KDIELNEVNA	NVLFNKNTTY	1260
	TATGNINGHV	DFQGNAGVIN	LNDIEIDGS	VTSTGNVNGT	LNFGSGKVT	GLINNIVMLQ	1320
	AGAGDVLSLA	SGNYSITEIQ	GNGNNNLTF	ANSHLTTDIN	KTGGQDLNLV	FINGGSSVGS	1380
	IGANAAGVDI	IINAGSVNFS	NTLKSGNIVI	SDGATMQVNN	NVTATDISGK	NANNGLTKLN	1440
	NHTPINITST	IGNNNAIGTI	EVANNDVTIT	GTLQAQNIHF	SNATQAATLT	LGAASQVTNI	1500
15	TTAGNNIHTL	EVTDFDGTND	GIIGDANNRL	KSIELTGNGT	VTINSPHVYS	SITTANNAQG	1560
	NVKLNIEGGI	TYDLGSKIKS	LANVQISED	TIRGDVYSKY	LNIDAGKTIN	PDRGDNNMNP	1620
	KNLDIPDALI	DLDVDLPRSL	LFNYFTDIIK	DNLNFPADDTA	TANFKDAVVI	DAHIDNGGIL	1680
	KFNDNAWLTQ	EIKNANIEI	ASDKFMLLQK	NIKAATLIAD	NANLVLLDNV	EVNTNLNVRD	1740
	IVLDLANYEL	KYTGNVTHNG	LLTIITYFDT	ALQKGGHILV	SQGSNVDMSD	LDNLIIKKA	1800
20	HSDITNITSD	TKHQIVKLET	GAITYPVPQT	KVIIDASEEQ	NKFVKWVADA	NGLVLLTDTG	1860
	GRDDTGGRDD	TRGRGNTDNG	CRDNCVGN	SNSSNEAGG	SSSDKNYGIT	DVVPIDFDPSP	1920
	ILDYTKNNYV	ASGIANQLIN	HVKDFGNTTD	AGKLLNDLGF	MSPNRVTETL	DRLSNRINVN	1980
	GLNEGVVGLN	GIEVENFLTD	IATNMDNFTA	KEIGNRLEEL	SDANTVNGLN	KTNTLLNNKI	2040
	NLKRNLNTNNQ	AIIAAGDEDN	IVTGIWGMSE	YGKIKQNSKN	SASGYQSNTG	GGIIGFDYNI	2100
25	DNSIVIGAAY	TMADSKVKHK	NDKNGDRTKA	KSNYSIYGL	YNWLTNNFFV	BAIGVYGRNK	2160
	IKNYEKRIIT	ITDQIAIGKF	INTFYSYELL	GGYNYLISHR	TTITPMFGMR	YATFKNNGYK	2220
	ENNTTFQNL	IKKNYYDKFE	TILGLNSVTH	YLSQDIIKIP	ELHWFINYQC	KNKLPNIDAR	2280
	LDGIDEPLTT	IRFKPAKIT	NLGGGISTKN	NMIEFGIRYN	LSLAKKYTAH	QGSCLKIKVNL	2340
30	<212> Type : PRT						
	<211> Length : 2340						
	SequenceName : SEQ ID 171						
	SequenceDescription :						
35	Sequence						

	<213> OrganismName : Rickettsia prowazekii						
	<400> PreSequenceString :						
40	MAQKPNFLKK	IISAGLVTAS	TATIVAGFSG	VAMGAAMQYN	RTTNAAATTF	DGIGFDQAAG	60
	ANIPVAPNSV	ITANANNPIT	FNTPNGHLNS	LFLDTANDLA	VTINEDTTLG	FITNIAQQAQ	120
	FFNFIVAAGK	ILNITGQGIT	VQEASNTINA	QNALTKVHGG	AAINANDLSG	LGSITFAAAP	180
	SVLEPNLINP	TTQEAPLTG	ANSKIVNGGN	GTLNITNGFI	QVSDNTFAGI	KTINIDDCQG	240
	LMFNSTPDAA	NTLNQLQVGN	TINFNGIDGT	GKLVLVSKNG	AATEFNVGTG	LGGNLKGII	300
	LNTAAVAGKL	ISQGAANAV	IGTDNGAGRA	AGFIVSVDN	NAATISGQVY	AKNMVIQSAN	360
45	AGGQVTFEHI	VDVGLGGTTN	FKTADSKVII	TENSNGGSTN	FGNLDTQIVV	PDKILKGNF	420
	IGDVKNNGNT	AGVITFNANG	ALVSASTDPN	IAVTNINAIE	AEGAGVVELS	GIHIAELRLG	480
	NGGSIFKLAD	GTIVNGPVNQ	NALMNNNALA	AGSIQLDGSA	IITGDIGNGG	VNAALQHITL	540
	ANDASKILAL	DGANIIGANV	GGAIHFQANG	GTIKLTNTQN	NIVVNFDDI	TTDKTGVVDA	600
	SSLTNNQTLT	INGSIGTVVA	NTKTLAQLNI	GSSKTILNAG	DVAINELVIE	NNGSVQLNHN	660
50	TYLITKTINA	ANQGQIIVAA	DPLNTNTTLA	DGTNLGSAEN	PLSTIHFATK	AANADSILNV	720
	GKGVNLYANN	ITNDANVGS	LHFRSGGTSI	VSGTVGGQQG	HKLNNLILDN	GTIVKFLGDT	780
	TFNGGTKIEG	KSILQISNNY	TTDHVESADN	TGTTLEFVNTD	PITVTLNKQG	AYFGVLKQVI	840
	ISGPGNIVFN	EIGNGIVHGG	IAANSISFEN	ASLGTSLFLP	SGTPLDVLT	KSTVNGTVD	900
	NFNAPIVVVS	GIDSMINNGQ	IIGDKKNIIA	LSLGSNDSIT	VNANTLYSGI	RTTKNNQGT	960
55	TLSGGMPNNP	GTIYGLGLEN	GSPKLKQVTF	TTDYNLNGSI	IANNVTINDY	VTLTTGGIAG	1020
	TDFDAKITLG	SVNGNANVRF	VDSTFSDPRS	MIVATQANKG	TVTYLGNALV	SNIGSLDTPV	1080
	ASVRFTGNDS	GAGLQGNIS	QNIDFGTYNL	TILNSNVILG	GGTTAINGEI	DLTNNLIFA	1140
	NGTSTWGDNT	SISTTLNVSS	GNIGQVVAIE	DAQVNATTG	TTTTIKIQDNA	NANFSGTQAY	1200
	TLIQGGARFN	GTLGAPNFAV	TGSNIFVKYE	LIRDSNQDYV	LTRTNDVLNV	VTTAVGNSAI	1260
60	ANAPGVSONI	SRCLESTNTA	AYNNMLLAKD	PSDVATFVGA	IATDTSAAVT	TVNLNDTQKT	1320
	QDLLSNRLGT	LRYLSNAETS	DVAGSATGAV	SSGDEAEVSY	GVWAKPFYNI	AEQDKKGGIA	1380
	GKAKKTGTVV	VGLDTLASDN	LMIGAAIGIT	KTDIKHQDYK	KGDKTIDINGL	SFSLYGSQQL	1440
	VKNFFAQGNA	IFTLNKVKSK	SQRYFFESNG	KMSKQIAAGN	YDNMTFGGNL	IFGYDYNAMP	1500
	NVLVTPMAGL	SYLKSSNENY	KETGTTVANK	RINSKFSDRV	DLIVGAKVAG	STVNITDIVI	1560
65	YPEIHSFVVH	KVNGKLSNSQ	SMLDGQTAPF	ISQPDRTAKT	SYNIGLSANI	KSDAKMEYGI	1620
	GYDFNSASKY	TAHQGTLLKVR	VNF				1643
	<212> Type : PRT						

<211> Length : 1643
SequenceName : SEQ ID 172
SequenceDescription :

5 Sequence

<213> OrganismName : Porphyromonas gingivalis W83

<400> PreSequenceString :

MARIILEAH	DVWEDGTGYQM	LWDADHNQYG	ASIPESSEFWF	ANGTIPAGLY	DPFEYKVPVN	60	
ADASFSP	TNFDVLDGTASADI	PAGTYDYVII	NPNPGIIV	GEVSKGNDY	VVEAGKTYHF	120	
TVQRQGP	DAASVVVTGEGG	NEFAPVQNLQ	WSVSGQTVTL	TWQAPASDKR	TYVLNESFDT	180	
QTLPNGW	TMI DADGDGHNWL	STINVYNTAT	HTGDGAMFSK	SWTASSGAKI	DLSPDNYLVT	240	
PKFTVP	ENGKLSYVWSSQEP	WTNEHYGVFL	STTGNEAANF	TIKLEETLG	SGKPAPMNLV	300	
KSEGVKA	PAPYQERTIDL	LSA YAGQQVYLAF	RHFVCTGIFR	LYLDDVAVSG	EGSSNDYTYT	360	
VYRDNV	VIAQLNLTATTFNQE	NVAPGQYNYC	VEVKYTAGVS	PKVCKDVTVE	GSNEFAPVQN	420	
LTGSAVG	QKVTWKWDAPNGT	PNPNGPTTTL	SESFENGIPA	SWKTIDADGD	GNNWTTTPPP	480	
GGSSFAG	HNSAICVSSASYI	NFEGPQNP	PDN YLVTPELSLP	NGGTLTFWVC	AQDANYASEH	540	
YAVYAS	STGN DASNANALL	EEVLTAKTVV	TAPEAIRGTR	VQGTWYQKTV	QLPAGTKYVA	600	
FRHFGCT	DFFWINLDDVEIK	ANGKRADTFE	TFESSTHGEA	PAEWTTIDAD	GDGQGWLC	660	
SGQLGWL	TAHGGTNVVASFS	WNGMALNPDN	YLISKDV	TGA TKVKYAVN	DGFPDGHYAV	720	
MISKTG	TNAG DFTVFEETP	NGINKGARF	GLSTEANGAK	PQSVWERTV	DLPAGTKYVA	780	
FRHYN	SDLN YILLDDIQFT	MGGSPTPTDY	TYTVYRDG	TK	EGLTETTF	EEDGVATGNH	840
EYCV	EYKYTA GVSPEKCVNV	TVDPVQFNPV	QNLGSAVGQ	KVTLKWDAPN	GTPNPNGT	900	
TLSE	SFENGI PASWKTIDAD	GDGNNWTTTP	PPGGTSFAGH	NSAICVSSAS	YINFEGPQNP	960	
DNYL	VTPELS LPNGGTLTFW	VCAQDANYAS	EHYAVYASST	GNDASN	FANA LLEEVLTAKT	1020	
VVTAP	EAIRG TRVQGTWYQK	TVQLPAGTKY	VAFRHF	GCTD FFWINLDDVE	IKANGKRADF	1080	
TET	FESSTHG EAPAEWTTID	ADGDGQGWLC	LSSGQLD	WLT AHGGTNVVAS	FSWNGMALNP	1140	
DNYL	ISKDV GATVKVYYA	VNDGFP	GDHY AVMISK	TGTM AGDFTVVFEE	TPNGINKGGA	1200	
RFGL	STEANG AKPQSVWIER	TVDL	PAGTKY VAFRHYNCSD	LNYILLDDIQ	FTMGGSP	TPT 1260	
DYTY	TVYRDG TKIKEGLTET	TFEEDGVATG	NHEVCVEVKY	TAGVSPKECV	NVTVPVQFN	1320	
PVQN	LTSAGV GQKVTWKDA	PNGTPNP	PNPG TTTLSE	SFEN GIPASWKTID	ADGDGNNWTT	1380	
TPPP	GGTSFA GHNSAICVSS	ASYINFEGPQ	NPDN	YLVTP ELSLPNGGTLT	FWVCAQDANY	1440	
ASEH	YAVYAS STGNDASNFA	NALLEEVLTA	KTVVTAP	EAI RGT	RVQGTWY QKT	VQLPAGT 1500	
KYVA	FRHFGC TRDFWINLDD	VEIKANGKRA	DFTET	FESST HGEAPAEWTT	IDADGDGQGW	1560	
LCLSS	GQLGW LTAHGGTNVV	ASFSWNGMAL	NPDN	YLSK DVTGATVKVYY	YAVNDGFP	PGD 1620	
HYAV	MISKTG TNAGDFTVVF	EETPN	GINKG GARFGL	STEA NGAKPQSVWI	ERTVDL	PAGT 1680	
KYVA	FRHYNC SDLN	YILLDD IQFTM	GGSP	PTDYTYTVYR	DG	TKIKEGLT ETTFEEDGVA 1740	
TGN	HEYCV	EV KYTAGVSPKE	CVNVTINPTQ	FNPVQNL	TAE QAPNSMDAIL	KWNAPASKRA 1800	
EVL	NEDFENG IPASWKTIDA	DGDGNNWTTT	PPPGSSFAG	HNSAICVSSA	SYINFEGPQNP	1860	
PDNY	LVTP EL SLPGGGTLTF	WVCAQDANYA	SEHYAVYASS	TGNDASN	FAN ALLEEVLTAK	1920	
TVVT	AP	EAIR GTRVQGTWYQ	KTVQLPAGTK	YVAFRHF	GCT DFFWINLDDV	VITSGNAPS	SY 1980
TYTY	IRNNTQ IASGV	TETTY RDPDLATGFY	TYGVKVVPY	PN GESAIETATL	NITSLADVTA	2040	
QKPY	TLTVVG KTTITVTCQGE	AMIDMNGRR	LAAGRNTVYV	TAQGGHYAVM	VVVDGKSYVE	2100	
KLAVK						2105	

45 <212> Type : PRT
<211> Length : 2105
SequenceName : SEQ ID 173
SequenceDescription :

50 Sequence

<213> OrganismName : Porphyromonas gingivalis W83

<400> PreSequenceString :

MKTSERIL	SY FFLCAVFS	L GSCEGLYAQV	TFPNYSPTAA	SSIAVCSGEE	TLIIDFTVVQ	60
EDSNGIK	VNV KLADGVEYV	V GTAVVSVTQG	NAVTV	AE TNV SNPNEPVFTV	KSADGNNVVE	120
LGTIVKL	TIK RRAVCTAWSN	AINAAETGFV	FKDKVTVTIG	DHSDSKESNS	YSVNYPNLTI	180
KQPAPQ	VNKQ IGETIVREFS	ITNGSQNP	TQ TVYLSIEY	PD EAYLTGVGAM	TLQAKLGASG	240
TYADLT	PVT NGKVRIY	TLS GSSLGPDHLL	TNGEIIYLKE	TFKLKTCAPV	TVYRVGWGCS	300
IDSQCE	IKTT AATITMAAGA	ANITGYSVTG	PDYRSPTFSL	CQPFELTIKF	SNSGAGGSMG	360
AAFNI	NTIGR NDYYRPRGFV	LHEFIDVKVN	GKPV	TNFKTD GSELDL	RFDG QFTEDPDGPG	420
VGLDD	VDGDG FYDDL	PVGAT ITITVT	VR LK CDQFTAC	NNA PNDLSDRGLI	LKTLYQTS	CD 480
RTSWID	PNTW RLSS	THLYL SRESVQDASH	MPTVIEK	DTP FDLKIM	TSYY SILSSYNNIW	540
YANPN	TRYV EIVFPQGM	TM PPKSDIEWTN	IKNHPIDGSL	VFTPPINL	PD ANITTSGNTM	600
TIVSP	SQERG FVTLHGV	KYD CTNNHEMVVE	YKIREVFN	YL HFPDCL	CPVG PIMCNTAKRY	660
VLGCD	PPCGR GAETSV	PKIE RADNSLG	WTD YTMRTQ	SRS NISAYDLAKA	LYMDEVNITA	720
TSIQHG	TASS LGARFV	LATG VDRVETL	TPL SADIKIF	RDG VQIVSV	DGYT TFRSIRNNN	780
AEQV	IDWDF	T SILPAGLLD	RDKVDV	VTRY RVTSQNAHRV	DTQVGREWFF	YNSTANVSPI 840

	WDEANPLTCL	ILVPEIYIMG	TFVVNGTDPH	VISQCTPTDL	GRVANHYARR	FGSGAFEYAN	900
	EYRPGVKIRN	IYLVKPKSYT	LNKVEYSNHR	NHSSLGTTMP	FEEINHTDVT	SQGEYNIYKY	960
	QLADNEKAHF	NITVKNAIGA	ALKVNVSPIC	ASSAVATNYD	KISYYVDYID	YYYYAATQPT	1020
	VPNSLDIVAD	QSAGSNGIYS	VSALNVYNRP	ILYTNKPSIA	LVNQSSEVEL	VGKTGEWKLR	1080
5	ISNPSSATAP	YVWLALPTTS	GLTIEKVDA	AGTEMAFTTY	SGGKMYRLSE	AGVPVGSALD	1140
	YTIHFTYSGC	SPIALKAMGG	WNCAYPLSL	DEVVCSQVI	DLKLKPLPAA	MELTEIAVPD	1200
	PTAAATLCST	LEYIYSIQST	DNANVYSPTF	SIFPEEGLVV	TPNQVQVEYP	AGSCNWAALN	1260
	VVMNSVNLQ	HPALTITIGYL	KGLKEGESND	NQRKILVKFY	IKTECSFVSG	KNFRVRADGR	1320
	NACNQNAKGS	GLAISTPPIR	INGAIEPYTT	SASTQLVTTT	TSQSDCKAPK	RVKVVQTVVG	1380
10	GETTPKAYLE	ITLPLGFKYV	TGSYAPDNTH	PGGVNASPAG	TEEVTLTANG	EDKIKINVKA	1440
	GLTSGQSFAY	TLEMKEDDDN	VPACGNHTIE	IVNVEEIEGL	WCEGVQCAET	LVVTGANKFE	1500
	FELDKPYLDI	TVISAVSTFS	GGKENLTIEY	KVSNTSTTQP	LKPGAVVTLF	SDKDNNQVFS	1560
	GGDVAVATQE	LVAEITNTTP	LTQIMKVKGV	SSSHTGNLVL	TILPKDGCYC	EIKSPMVTLN	1620
	HTGSNYSGAN	TVGKPNWEKE	PNNWTNDQVP	DAEDEVFAT	EVNNPTDPNN	PKSGPAKENL	1680
15	HLDDIHQNGT	AGRVIGNLIN	DSKDLVITT	GNQLTINGVV	EDNNPNVGTI	VVKSSKDNPT	1740
	GTLLFANPGN	NQNVGTVFVF	YNQGYDCADC	GMYYRSWQYF	GIPVNESDFP	YDHDVGNATV	1800
	NQWVEPFNGD	KWRPAPYAPD	TKLQKFKGYQ	ITNDVQAQPT	GVYSFKGTLC	VCDAPLNLTR	1860
	TSGVNYSGAN	LIGNSYTGAI	DIKQGIIVFP	EVEQTVYLFN	TGTRDQWRKL	NGSTVSGYRA	1920
	GQYLSVPKNT	AGQDNLPDRI	PSMHSFLVKM	QNGASCTLQI	LYDKLLKNTT	VNNGNGTQIT	1980
20	WRSGNSGSAN	MPSLVMDVLG	NESADRLWIF	TDGGLSFGFD	NGWDGRKLTE	KGLSGLYAMS	2040
	DIGNDKFQVA	GVPELNNLLI	GFDADKDGQY	TLEFALSDFH	AKGGVFLEDL	SRGVTRRVVD	2100
	GGYSYFDAQR	GDGSGARFRL	YDEEWVESAE	VSVLVGTACK	RIVITNNSEH	ACQANVYTTD	2160
	GKLLIRLDVK	PGSKSMTEPL	VDGVYVVSLO	SPATSSNVRK	VVVN		2204
	<212> Type : PRT						
25	<211> Length : 2204						
	SequenceName : SEQ ID 174						
	SequenceDescription :						
	Sequence						
30	-----						
	<213> OrganismName : Porphyromonas gingivalis W83						
	<400> PreSequenceString :						
	MNKFKYKSLQ	SGLAASFVMA	TALTASQAIS	PGGEPLSFSS	RSAGTHSFDD	AMTIRLTPDF	60
	NPEDLIAQSR	WQSQRDGRPV	RIGQVIPVDV	DFASKASHIS	SIGDVDVYRL	QFKLEGAKAI	120
35	TLYYDAFNIP	EGGRLYIYTP	DHEIVLGAYT	NATHRRNGAF	ATEPVPGSEL	IMDYEVSRGG	180
	TLDPDIKISGA	GYIFDKVGGP	PVTDNHYGIG	EDDSDSDECI	NINCEGADW	QAEKNGVVMQ	240
	IMVKGQYISM	CSGNLLNNTK	GDFTPLIISA	GHCASITTNF	GVTQSELDKW	IFTFHYEKRQ	300
	CSNGTLAIFR	GNSIIGASMK	AFLPIKGSKD	GLLQLNDEV	PLRYRVYYNG	WDSTPDIPSS	360
	GAGIHHAPAGD	AMKISILKKT	PALNTWISSS	GSGGTDDHFI	FKYDQGGTEG	GSSGSSSLFNQ	420
40	NKHVVGTLTG	GAGNCGGTEF	YGRNLNHWNE	YASDGNTRM	DIYLDPQNNQ	QTTILNGTYR	480
	DGYKPLPSVP	RLLLQSTGDQ	VELNWTAVPA	DQYPPSSQVE	YHIFRNGKEI	ATTKELSYSD	540
	AIDESTIAGS	IIRYEVSAEF	IYPSPLDGVE	SYKDTDKTSA	DLAIGDIQTK	LKPDVTPPLG	600
	GGVSLSWKVP	FLSQLVSRFG	ESPMPVFKTF	EVPIVSAAAA	QTPNPPVGVV	IADKFMAGTY	660
	PEKAAIAAVY	VMPSAPDSTF	HLFLKSNTRN	RLQKVTTTSD	WQAGTWLRIN	LDKPPFPVND	720
45	HMLFAGIRMP	NKYKLNRAIR	YVRNPDNLFS	ITGKKISYNN	GVSFEGYGIP	SLLGMYMAIKY	780
	LVVNTDAPKI	DMSLVQEPYA	KGTNVAPFPE	LVGIYVYKNG	TFIGTQDPSV	TTYSVSDGTE	840
	SDEYEIKLVY	KSGSISNGVA	QIENNNAVVA	YPSVVTDRFS	IKNAHMHVAA	ALYSLDGGKQV	900
	RSWNNLRNGV	TFSVQGLTAG	TYMLVMQTAN	GPVSQKIVKQ			940
	<212> Type : PRT						
50	<211> Length : 940						
	SequenceName : SEQ ID 175						
	SequenceDescription :						
	Sequence						
55	-----						
	<213> OrganismName : Porphyromonas gingivalis W83						
	<400> PreSequenceString :						
	MKNLNKFVSI	ALCSSLGGM	AFAQQOTELGR	NPNVRLLEST	QQSVTKVQFR	MDNLKFTEVQ	60
	TPKGMQAVPT	YTEGVNLSEK	GMPTLPILSR	SLAVSDTREM	KVEVVSSKFI	EKKNVLIAPS	120
60	KGMIMRNEDP	KKIPYVYGKS	YSQNKFFPGE	IATLDDPFIL	RDVRGQVNVF	APLQYNPVTK	180
	TLRIYTEITV	AVSETSEQGK	NILNKKGTFA	GFEDTYKRMF	MNYEPGRYTP	VEEKQNGRMI	240
	VIVAKYEGD	IKDFVDWKNQ	RGLRTEVKVA	EDIASPVTAN	AIQQFVKQEV	EKEGNDLTYV	300
	LLIGDHKDIP	AKITPGIKSD	QVYQIVGND	HYNEVFIGRF	SCESKEDLKT	QIDRTIHYER	360
	NITTEDKWLQ	QALCIASAEG	GPSADNGESD	IQHENVIANL	LTQYGYTKII	KCYDPGVTPK	420
65	NIIDAFNGGI	SLANYTGHGS	ETAWGTSHFG	TTHVKQLTNS	NQLPFIQDVA	CVNGDFLFMS	480
	PCFAEALMRA	QKDGKPTGTV	AIIASTINQS	WASPMRGQDE	MNEILCEKHP	NNIKRTFGGV	540
	TMNGMFAMVE	KYKKGDEKML	DTWTVFGDPS	LLVRTLVPFK	MQVTAPAQIN	LTDASVNVSC	600

	DYNGAIATIS	ANGKMFSGSAV	VENGATATINL	TGLTNESTLT	LTVVGYNKET	VIKTINTNGE	660
	PNPYQVPSNL	TATTQGGQVKT	LKWDAPSTKT	NATTNTARSV	DGIRELVLLS	VSDAPELLRS	720
	GQAEIVLEAH	DVWNDGSGYQ	ILLDADHDQY	GQVIPSdTHT	LWPNCSSVPAN	LFAPFEYTVP	780
	ENADPSCSPT	NMIMDGTASV	NIPAGTYDFA	IAAPQANAKI	WIAGQGPTKE	DDYVFEAGKK	840
5	YHFLMKKMG	GDGTELTISE	GGGSDYTYTV	YRDGTIKIEG	LTATTFEEDG	VAAGNHEYCV	900
	EVKYTAGVSP	KVCKDVTVEG	SNEFAPVQNL	TGSAVGQKVT	LKWDAPNGTP	NPNPNPNPNP	960
	NPGTITLSES	FENGIPASWK	TIDADGDGHG	WKPGNAPGIA	GYNSNGCVYS	ESFGLGGIGV	1020
	LTPDNYLITP	ALDLPNGGKL	TFWVCAQDAN	YASEHYAVYA	SSTGNDASNF	TNALLEETIT	1080
	AKGVRSPETI	GRIGQGTWRQ	KTVDLPAGTK	YVAFRHFQST	DMFYIDLDEV	EIKANGKRAD	1140
10	FTETFESSHT	GEAPAEWTTI	DADGDGQGWL	CLSSGQLDWL	TAHGGTNNVS	SFSWNGMALN	1200
	PDNYLISKDV	TGATKVKYVY	AVNDGFGPDH	YAVMTSKTGT	NAGDFTVVFE	ETPNGINKGG	1260
	ARFGLSTEAD	GAKPQSVWIE	RTVDLPAGTK	YVAFRHYNCS	DLNYILLDDI	QFTMGGSPTP	1320
	TDYTYTVYRD	GTRIKIEGLTE	TTFEEDGVAT	GNHEYCVEVK	YTAGVSPKKC	VNVTVNSTQF	1380
	NPVKNLKAQP	DGGDVVLKWE	APSAKKTEGS	REVKRIGDGL	FVTIEPANDV	RANEAQVVL	1440
15	ADNVWGNTG	YQFLLDADHN	TFGSVIPATG	PLFTGTASSD	LYSANFEYLI	PANADPVVTT	1500
	QNIIVTGQGE	VVIPGGVYDY	CITNPEPASG	KMWIAGDGGN	QPARYDDFTF	EAGKKYTFMT	1560
	RRAGMGDGT	MEVEDDSPAS	YTYTVYRDGT	KIKEGLTETT	YRDAGMSAQ	HEYCVEVKYT	1620
	AGVSPKVCVD	YIPDGVADVT	AQKPYTLTVV	GKTTITVTCQG	EAMIYDMNGR	RLAAGRNTVV	1680
	YTAQGGYYAV	MVVVDGKSYV	EKLAIK				1706
20	<212> Type : PRT						
	<211> Length : 1706						
	SequenceName : SEQ ID 176						
	SequenceDescription :						
25	Sequence						

	<213> OrganismName : Porphyromonas gingivalis W83						
	<400> PreSequenceString :						
	MKRKPLFSAL	VILSGFFGSV	HPASAQKVPA	PVDGERIIME	LSEADVECTI	KIEAEDGYAN	60
30	DIWADLNGNG	KYDSGERLDS	GEFRDVEFRQ	TKAIVYGKMA	KFLFRGSSAG	DYGATFIDIS	120
	NCTGLTAFDC	FANLLTELDEL	SKANGLTFVN	CGKNQLTKLD	LPANADIETL	NCSKNKITS	180
	NLSTYTKLKE	LYVGDNGLTA	LDLSANTLLE	ELVYSNNEVT	TINLSANTNL	KSLYCINNKM	240
	TGLDVAANKE	LKILHCNNNQ	LTALNLSANT	KLTTLSEFFNN	ELTNIDLSDN	TALEWLFCNG	300
	NKLTKLVDVA	NANLIALQCS	NNQLTALDLS	KTPKLTTLNC	YSNRIKDTAM	RALIESLPTI	360
35	TEGEGREFVPY	NDEGEGEEEN	VCTTEHVEMA	KAKNWKVLTS	WGEPPPGITA	LISIEGESEY	420
	SVYAQDGILY	LSGMEQGLPV	QVYTVGGSM	YSSVASGSAM	EIQLPRAAY	VVRIGSHAIK	480
	TAMP						484
	<212> Type : PRT						
	<211> Length : 484						
40	SequenceName : SEQ ID 177						
	SequenceDescription :						
	Sequence						

45	<213> OrganismName : Shigella flexneri 2a str. 2457T						
	<400> PreSequenceString :						
	MKRAITLFAV	LLMGWSVNAW	SFACKTANGT	AIPIGGGSAN	VYVNLAPVVN	VGQNLVVDLS	60
	TQIFCHNDYP	ETITDYVTLO	RGSAYGGVLS	NFSGTVKYS	SSYPFFTSE	TPRVVYNSRT	120
	DKPWPVALYL	TPVSSAGGVA	IKAGSLIAVL	ILRQTNNYNS	DDFQFVWNIY	ANNDVVVPTG	180
50	GCDVSARDVT	VTLPDYPGSV	PIPLTVYCAK	SQNLGYLSSG	TTADAGNSIF	TNTASFSPAQ	240
	GVGVQLTRNG	TIIPANNTVS	LGAVGTSAVS	LGLTANYART	GGQVTAGNVQ	SIIGVTFVYQ	300
	<212> Type : PRT						
	<211> Length : 300						
55	SequenceName : SEQ ID 178						
	SequenceDescription :						
	Sequence						

60	<213> OrganismName : Shigella flexneri 2a str. 2457T						
	<400> PreSequenceString :						
	MGIKQHNGNT	KADRLAELKI	RSPSIQLIKF	GAIGLNAIIF	SPLLIAADTG	SQYGTNITIN	60
	DGDRITGDTA	DPSGNLYGVM	TPAGNTPGNI	NLGNDVTNVN	NDASGYAKGI	IIQKNSSLT	120
	ANRLTVDVVG	QTSAGINLI	GDYTHADLGT	GSTIKSNDG	IIIGHSSTLT	ATQFTIENS	180
65	GIGLTIIDYG	TSVDLGS	IKTDGSTGVY	IGGLNGNNAN	GAARFTATDL	TIDVQGSAM	240
	GINVQNSV	DLGTNSTIKT	NGDNAHGLWS	FGQVSANALT	VDVTGAAANG	VEVRGGTTTI	300
	GADSHISSAQ	GGGLVTSSSD	ATINFSGTAA	QRNSIFS	YGASQATATA	VINMQNTDIT	360

VDRNGSLALG LWALSGGRIT GDSLAIITGAA GARGIYAMTN SQIDLTSDLV IDMSTPDQMA 420
IATQHDDGYA ASRINASGRM LINGSVLSKG GLINLDMHPG SVWTGSSLSD NVNGGKLDVA 480
MNNSVWNVTS NSNLDTLALS HSTVDFASHG STAGTFTTLN VENLSGNSTF IMRADVVGE 540
NGVNNRGDGL NISGSSAGNH VLAI RNQSGE ATTGNEVLTV VKTTDGAASF SASSQVELGG 600
5 YLYDVRKNGT NWELYASGTV PEPTNPPEPT PAPAQPPIVN PDPTPEPAPT PKPTTTADAG 660
GNYLNVGYLL NYVENRTLMO RMGDLRNQSK DGNIWLRSYG GSLSDFASGK LSGFDMGYSG 720
IQFGGDKRLS DVMPLYVGLY IDSTHASPDY SGGDGTARSD YMGMYASYMA QNGFYSDLVI 780
KASRQKNSFH VLDSQNNGVN ANGTANGMSI SLEAGQRFNL SPTGYGFYIE PQTQLTYSHQ 840
NEMAMKASNG LNIHLNHYES LLGRASMILG YDITAGNSQL NVYVKTGAIR EFSGDTEYLL 900
10 NDSREKYSFK GNGWNNNGVGV SAQYNKQHTF YLEADYTQGN LFDQKQVNGG YRFSF 955

<212> Type : PRT
<211> Length : 955
SequenceName : SEQ ID 179
15 SequenceDescription :

Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
20 <400> PreSequenceString :
MSKFVKTAIA AAMVMGVFTS TATIAAGNNG TARFYGTIED SVCSIVPDDH KLEVDMGDIG 60
AEKLNKNGTT TPKSFQIRLQ DCVFDTQETM TTTFTGTVSS ANSGNYTIF NTDTGAAFN 120
VSLAIGDSLQ TSYKSGMGID QKIVKDTSTN KKGAKQTLNF NAWLVGAADA PDLGNFEANT 180
TFQITYL 187

25 <212> Type : PRT
<211> Length : 187
SequenceName : SEQ ID 180
SequenceDescription :

30 Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MKIKTLAIVV LSALSLSAA ALADTTTVNG GTIHFKEGVV NAACAVDAGS VDQTVQLGQV 60
35 RTASLKQAGA TSSAVGFNIQ LNDCTTTVAT KAAVAFLGTA IDATRTDVLA LQSSAAGSAT 120
NVGVQILDRT GNALTLDGAT FSAQTTLNNG TNTIPFQARY YAIGETPGA ANADATFKVQ 180
YQ 182

<212> Type : PRT
<211> Length : 182
40 SequenceName : SEQ ID 181
SequenceDescription :

Sequence

45 <213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MASISLGVG SGLDLSSILD SLTAAQKATL TPISNQSSF TAKLSAYGTL KSALTTFQTA 60
NTALSKADLF SATSTTSSTT AFSATTAGNA IAGKYTISVT HLAQAQTLTT RTTRDDTKTA 120
IATSDSKLTI QQGDDKDPIT IDISAANSSL SGIRDAINNA KAGVSASIIN VGNGEYRLSV 180
50 TSNDTGLDNA MTLVSVGDDA LQSFMGYDAS ASSNGMEVSV AAQNAQLTVN NVAIENSSNT 240
ISDALENITL NLNDVTTGNQ TLTITQDTSK VQTAIKDWNV AYNSLIDTFS SLTKYTAVDA 300
GADSQSSSNG ALLGDSTLRT IQTQLKSMLS NTVSSSSYKT LAQIGITTD P SDGKLELDAD 360
KLTAALKKDA SGVGALIVGD GKKTGITTTI GSNLTSWLST TGIKAATDG VSKTLNKLTK 420
DYNAASDRID AQVARYKEQF TQLDVLMTSL NSTSSYLTTQ FENNSNSK 468

55 <212> Type : PRT
<211> Length : 468
SequenceName : SEQ ID 182
SequenceDescription :

60 Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MEGKADNVVL ENGGRLDVLT GHTATNTRVD DGGTLDVRNG GTATTVSMGN GGVLLADSGA 60
65 AVSGTRSDGK AFSIGGGQAD ALMLEKGGSSF TLNAGDTATD TTVNGGLFTA RGGTLAGTTT 120
LNNGAILTSL GKTVMNDTLT IREGDALLQG GALTNGSVE KSGSGTLTVS NTTLTQKAVN 180
LNEGTLTLND STVTTDVIAQ RGTALKLTGS TVLNGAIDPT NVTLASGATW NIPDNATVQS 240

```

VVDLDSHAGQ IHFTSTRTGK FVPATLKVKV LNQGNGTISL RVRPDMAQNN ADRLVIDGGR      300
ATGKTILNLV NAGNSASGLA TSGKGIQVVE AINGATTEEG AFIQGNKLQA GAFNYSLNRD      360
SDESWYLRSE NAYRAEVPLY ASMLTQAMDY DRILAGSRSH QTGVSGENNS VRLSIQGGHL      420
GHDNNGGIAR GATPESGSGY GFVRLEGDLI RTEVAGMSVT AGVYGAAGHS SVDVKDDGDS      480
5  RAGTVRDDAG SLGGYLNLIH NASGLWADIV AQGTRHSMKA SSDNNDFRVR GWGWLGSLET      540
GLPFSITDNL MLEPQLQYTW QGLSLDDGQD NASYVKFGHG SAQHVRAGFR LGSHHDMNFG      600
KGTSSRDTLR GSAKHSVREL PVNWWVQPSV IRTFSSRGDM SMGTAAAGSN MTFSPSQNGT      660
SLDLQAGLEA RVRENTLGV QASYAHSING SSARGYNSQA TLNVTF      706
<212> Type : PRT
10 <211> Length : 706
    SequenceName : SEQ ID 183
    SequenceDescription :

Sequence
-----
15 <213> OrganismName : Shigella flexneri 2a str. 2457T
    <400> PreSequenceString :
MAFSQAVSGL NAAATNLDDVI GNNIANSATY GFKSGTASFA DMFAGSKVGL GVKVAGITQD      60
FTDGTNTINT RGLDVAISQN GFFRLVDSNG SVFYSRNGQF KLDENRNLLN TQGLQLTGYP      120
20 VTGTPPTIQQ GANPTNISIP NTLMAAKTTT TASMQINLNS SDPLPTVTPF SASNADSYNK      180
KGSVTVFDSQ GNAHDSVYF VKTGDNNWQV YTQDSSDPNS IAKTATTLEF NANGTLVDGA      240
MANNIATGAI NGAEPATFSL SFLNSMQONT GANNIVATTQ NGYKPGDLVS YQINDDGTVV      300
GNNSNEQTQL LGQIVLANFA NNEGLASEGD NVWSATQSSG VALLGTAGTG NFGTLTNGAL      360
EASNVDSKE LVNMIVAQRN YKSNAQTIKT QDQILNTRVN LR      402
25 <212> Type : PRT
    <211> Length : 402
    SequenceName : SEQ ID 184
    SequenceDescription :

Sequence
-----
30 <213> OrganismName : Shigella flexneri 2a str. 2457T
    <400> PreSequenceString :
MKLVHMASGL AVAIALACA DKSADIQTPA PAANTSISAT QQPAIQPNV SGTVWIRQKV      60
35 ALPPDAVLTV TLDASLADA PSKVLAKAV RTEGKQSPFS FVLFPNPADV QPNARILLSA      120
AITVNDKLVF ITDTVQPVIN QGGTKADLTL VPVQQTAVPV QASGGATTTV PSTSPTQVNP      180
SSAVPAPTQY      190
    <212> Type : PRT
    <211> Length : 190
    SequenceName : SEQ ID 185
    SequenceDescription :

Sequence
-----
45 <213> OrganismName : Shigella flexneri 2a str. 2457T
    <400> PreSequenceString :
MIIKSGGRW QLSLLASVVI SAFFLNTAYA WQGEYIVDTQ PGHSTERYTW DSDHQPDYND      60
ILSQRIQSSQ RALGLEVNLA EETPVDVTSS MSMGWNFPY EQVTGTPVAA LHYDGTTSMT      120
YNEFGDSTTT LTDPLWHASV SSLGWRVDSR LGDLRPWAI SYNQQFGENI WKAQSGLSRM      180
50 TATNQGNWL DVTVGADMLL NQNIAAYAAL TQAENTTNNS DYLYTMGVSA RF      232
    <212> Type : PRT
    <211> Length : 232
    SequenceName : SEQ ID 186
    SequenceDescription :

Sequence
-----
60 <213> OrganismName : Shigella flexneri 2a str. 2457T
    <400> PreSequenceString :
MKWCKRGYVL AAMLALASAT IQAADVTITV NGKVVAKPCT VSTTNATVDL GDLYSFSLMS      60
AGAASAWHDV ALELTNCPVG TSRVTASFSG AADSTGYIKN QGTAQNIQLE LQDDSGNTLN      120
TGATKTQVD DSSQSAHFPL QVRALTVNGG ATQGTIQAVI SITYTYS      167
    <212> Type : PRT
65 <211> Length : 167
    SequenceName : SEQ ID 187
    SequenceDescription :

```

Sequence

5 <213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MKRAPLITGL LLISTSCAYA SSGGCGADST SGATNYSSVV DDVTVNQTDN VTGREFTSAT 60
LSSTNWQYAC SCSAGKAVKL VYMVSPVLTG TGHQTGYVKL NDSLDIKTTL QANDIPGLTT 120
DQVVSVNTRF TQIKSSTVYS AATQTGVCQG DTSRYGPVNI GANTTFTLYV TKPFLGSMTI 180
PKTDIAVIK AWVDGMSGPS TGDHFDLVKL SIQGNLTAPQ SCKINQGDVI KVNFGFINGQ 240
10 KFTTRNAMPD GFTPVDFDIT YDCGDTSKIK NSLQMRIDGT TGVVDQYNLV ARRRSSDNVP 300
DVGIRIENLG GGVANIPFQN GILPVDPSGH GTVNMRAWPV NLVGGELETG KFQGTATITV 360
MVR 363
<212> Type : PRT
<211> Length : 363
15 SequenceName : SEQ ID 188
SequenceDescription :

Sequence

20 <213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MQKNAAHTYA ISSLLVLSLT GCAWIPSTPL VQGATSAQPV PGPTPVANGS IFQSAQPINY 60
GYQPLFEDRR PRNIGDTLTI VLQENVSASK SSSANASRDG KTNFGFDTVP RYLQGLFGNA 120
RADVEASGGN TFNGKGGANA SNTFSGTLTV TVDQVLVNGN LHVVGKQIA INQGTETIRF 180
25 SGVVPNPTIS GSNTVPSTQV ADARIEYVGN GYINEAQNMG WLQRFFLNLS PM 232
<212> Type : PRT
<211> Length : 232
SequenceName : SEQ ID 189
30 SequenceDescription :

Sequence

35 <213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MKRHLNTCYR LVWNHITGAF VVASELARAQ GKRGGVAVAL SLAAVTSPLV LAADIVVHPG 60
ETVNGGTLVN HDNQFVSGTA DGVTVSTGLE LGPDSDENTG GQWIKAGGTG RNTTVTANGR 120
QIVQAGGTAS DTVIRDGGGQ SLNGLAVNTT LDNRGEQWVH GGGKAAGTII NQDGYQTIKH 180
GGLATGTIVN TGAEGGPESE NVSSGQMVGG TAESTTINKN GRQVIWSSGM ARDTLIYAGG 240
40 DQTVHGEAHN TRLEGGNQYV HNGGTATETL INRDGWQVIK EGGTAAHTTI NQKESCR 297
<212> Type : PRT
<211> Length : 297
SequenceName : SEQ ID 190
45 SequenceDescription :

Sequence

50 <213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MMMKTIKHLL CCAIAASALI STGVHAASWK DALSSAASEL GNQNSTTQEG GWSLASLTNL 60
LSSGNQALSA DNMNNAAGIL QYCAKQKLAS VTDAENIKNQ VLEKLGLNSE EQKEDTNYLD 120
GIQGLLKTGD GQQLNLDNIG TTPLAEKVKT KACDLVLKQG LNFIS 165
<212> Type : PRT
55 <211> Length : 165
SequenceName : SEQ ID 191
SequenceDescription :

Sequence

60 <213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MFKGQKTLAA LAVSLLFTAP VYAADEGSGE IHFKGEVIEA PCEIHQDDID KEVELGQVTT 60
SHINQSHHSD AVAVDLLLVN CDLENSNNGS GKGISKVAVT FDSSAKTTGA DPILNNTSTG 120
65 EATGVGVRLM NKDQSNIVLG TATPDIDLAP TSSEQTLNFF AWMEQIDQAT PVTPGAUTAN 180
ATYVLDYK 188
<212> Type : PRT

<211> Length : 188
SequenceName : SEQ ID 192
SequenceDescription :

5 Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MSAGSPKFTV RRIAALSLVS LWLAGCSDTS NPPAPVSSVN GNAPANTNSG MLITPPPKMG 60
TTSTAQQPQI QPVQQPQIQ TQPPQIQPVQ PVAQQPVQME NGRIVYNRQY GNIPKGSYSG 120
10 STYTVKKGDT LFYIAWITGN DFRDLAQRNN IQAPYALNVG QTLQVGNASG TPITGGNAIT 180
QADAAEQGVV IKPAQNSTVA VASQPTITYS ESSGGEQSANK MLPNNKPTAT TVTAPVTVPT 240
ASTTEPTVSS TSTSTPISTW RWPTEGKVIE TFGASEGGNK GIDIAGSKGQ AIIATADGRV 300
VYAGNALRGY GNLIILIKHND DYLSAYAHND TMLVREQQEV KAGQKIATMG STGTSSTRLH 360
15 FEIRYKGSV NPLRYLPQR 379

<212> Type : PRT
<211> Length : 379
SequenceName : SEQ ID 193
SequenceDescription :

20 Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MIKFLSALIL LLVTTAAQAE RIRDLSVQV VRQNSLIGYG LVVGLDGTGD QTTQTPFTTQ 60
25 TLNNMLSQLG ITVPTGTNMQ LKNVAAMVT ASLPPFGRQG QTIDVVVSSM GNAKSLRGGT 120
LLMTPLKGVD SQVYALAQGN ILVGAGASA GGSSVQVNQL NGGRITNGAV IERELPSQFG 180
VGNTLNLQLN DEDFSMAQQI ADTINRVRGY GSATALDART IQVRVPSGNS SQVRFLADIQ 240
NMQVNVTPQD AKVVINSRTG SVVMNREVTI DSCAVAQGNL SVTVNRQANV SQPDTPFGGG 300
30 QTVVTPQTQI DLRQSGGSLQ SVRSSASLNN VVRALNALGA TPMDLMSILQ SMQSAGCLRA 360
KLEII 365

<212> Type : PRT
<211> Length : 365
SequenceName : SEQ ID 194
SequenceDescription :

35 Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MKRSIIAAAV FSSFFMSAGV FAADVDTGTL TIKGNIAESP CKFEAGGDSV SINMPTVPTT 60
40 VFEKGAKYST YDDAVGVTSS MLKISCPKEV AGVKLSLITN DKITGNDKAI ASSNDFVGDN 120
SDVLDVSAPF NIESYKTAEG QYAIPFKAKY LKLTDNSVQS GDVLSLVMR VAQD 174

<212> Type : PRT
<211> Length : 174
SequenceName : SEQ ID 195
SequenceDescription :

50 Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MAVQKNVIK ILAGTFALML SGCVTVPDAI KGSSTTPQQD LVRVMSAPQL YVGQEARFGG 60
55 KVVAVQNQQG KTRLEIATVP LDGARGPTLG EPSRGRIYAD VNGFLDPVDF RGQLVTVVGP 120
ITGAVDGKIG NTPYKFMVMQ VTGYKRWHLT QQVIMPPQPI DPWFYGGRGW PYGYGGGWY 180
NPGPARVQTV VTE 193

<212> Type : PRT
<211> Length : 193
SequenceName : SEQ ID 196
SequenceDescription :

60 Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MRNKPFYLLC AFLWLAVSRV LAADSTITIR GYVRDNGCSV AAESTNFTVD LMENAAQFN 60

NIGATTPVVP FRILLSPCGN AVSAVKVGFT GVADSHNANL LALENTVSAA AGLGIQLLNE 120
QQNQIPLNAP SSAISWTTLT PGKPNTLNFY ARLMATQVPV TAGHINATAT FTLEYQ 176

<212> Type : PRT
5 <211> Length : 176
SequenceName : SEQ ID 197
SequenceDescription :

Sequence
10 -----
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MKKLTVAALA VTLLSGSAF AHEAGEFFMR AGSATVRPTE GAGGTLGSLG GFSVTNNTQL 60
GLTFTYMATD NIGVELLAAT PFRHKIGTRA TGDIA TVHHL PPTLMAQWYF GDASSKFRPY 120
15 VGAGINYTTF PDNGFNDHGK EAGLSDSLK DSWGAAGQVG VDYLINRDWL VNMSVWYMDI 180
DTTANYKLGG AQQHDSVRLD PWVFMFSAGY RF 212

<212> Type : PRT
<211> Length : 212
SequenceName : SEQ ID 198
20 SequenceDescription :

Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
25 <400> PreSequenceString :
MFFFRGKILS AGRNLKKS LG IVMFLSVGLL LAGCSGSKSS DTGTYSGSVY TVKRGDTLYR 60
ISRTTGTSVK ELARLNGISP PYTIEVGQKL KLGAKSSSS TRKSTAKSTT KTASVTPSSA 120
VPKSSWPPVG QRCWLWPTTG KVIMPYSTAD GGNKGIDISA PRGTPPIYAAG AGKV VYVGNQ 180
LRGYGNLIMI KHSEDIYITAY AHNDTMLVNN GQSVKAGQKI ATMGSTDAAS VRLHFQIRYR 240
30 ATAIDPLRYL PPQGSKPKC 259

<212> Type : PRT
<211> Length : 259
SequenceName : SEQ ID 199
SequenceDescription :

35 Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
40 MAQVINTNSL SLITQNNINK NQSALSSSIE RLSSGLRINS AKDDAAGQAI ANRFTSNIKG 60
LTQAARNAND GISVAQTTEG ALSEINNMLQ RIRELTVQAS TGNSDSDDL SIQDEIKSRL 120
DEIDRVSGQT QFNGVNV LAK DGSMKIQVGA NDGQTITIDL KKIDSDTLGL NGFNVNGGGA 180
VANTAASKAD LVAANATVVG NKYTVSAGYD AAKASDLLAG VSDGDTVQAT INNGFGTAAS 240
ATNYKYDSAS KSYSFDTTSA AADVQKYL T PGVGD TAKGT ITIDGSAQDV QISSDGKITA 300
45 SNGDKLYIDT TGR LTKNGSG ASLTEASLST LAANNTKATT IDIGGTSISF TGNSTTPDTI 360
TYSVTGAKVD QAAFDKAVST SGNNVDFTTA GYSVNGTTGA VTKGVDSVYV DNNEALTTSD 420
TVDFYLQDDG SVTNGSGKAV YKDADGKLTT DAETKAATTA DPLKALDEAI SSIDKFRSSL 480
GAVQNR LDSA VTNLNNTT N LSEAQSRIQD ADYATEVSNM SKAQIIQQAG NSVLAKANQV 540
PQQVLSLLQG 550

50 <212> Type : PRT
<211> Length : 550
SequenceName : SEQ ID 200
SequenceDescription :

55 Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
60 MKKIACLSAL AAVLAFTAGT SVAATSTVTG GYAQSDAQGQ MNKMGGFN LK YRYEEDNSPL 60
GVIGSFYTE KSR TASSGDY NKNQYYGITA GPAYRINDWA SIYGVVGVGY GKFTTEYPT 120
YKHDTSYDGF SYGAGLQFNP MENVALDFS Y EQSRIRSV DV GTWIAGVGYR F 171

<212> Type : PRT
<211> Length : 171
65 SequenceName : SEQ ID 201
SequenceDescription :

Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
5 MKRNIIGGAF TLASLMLAGH ALAEDGVVNF VGEIVDTTCE VTSDTADQIV PLGKVSNAF 60
SGVGSASPQ KFSIKLENC P ATYTQAAVRF DGTEAPGGDG DLKVGTPPLTA GNP GDFTGTG 120
QAIAATGVGI RIFNQSDNSQ VKLYNDSAYT AIDAEGKAEM KFIARYVATN ATVTAGTANA 180
DSQFTVEYKK 190
<212> Type : PRT
10 <211> Length : 190
SequenceName : SEQ ID 202
SequenceDescription :

Sequence

15 <213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MKKSTLALVV MGIVASASVQ AAEIYNK DGN KLDVYGK VKA MHYMSDNASK DGDQSYIRFG 60
FKGETQINDQ LTGYGRWEAE FAGNKAESDT AQKQTRLAFA GLKYKDLGSF DYGRNLGALY 120
20 DVEAWTDMFP EFGGDSSAQ T DNFMTKRASG LATYRNTDFF GVIDGLNLTL QYQGNENRD 180
VKKQNGDGF G TSLTYDFGGS DFAISGAYTN SDR TNEQNLQ SRGTGKRAEA WATGLKYDAN 240
NIYLATFYSE TRKMTPI TGG FANKTQNFEA VAQYQDFDGL RPSLGYVLSK GKDIEGIGDE 300
DLVNYIDVGA TYFFNKNMSA FVDYKINQLD SDNKLININD DTVAVGMTYQ F 351
<212> Type : PRT
<211> Length : 351
SequenceName : SEQ ID 203
SequenceDescription :

30 Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MRKQWL GICI AAGMLA ACTS DDGQQQT VSV PQPAVCNGPI VEISGADPRF EPLNATANQD 60
35 YQRDGKSYKI VQDPSRFSQA GLAAIYDAEP GSNLTASGEA FDP TKLTAAH PTLPIPSYAR 120
ITNLANGEMI VVRINDRGPY GNDRVISLSR AAADRLNTSN NT'KVRIDPII VAQDGSLSGP 180
GMACTTVAKQ TYALPAPPDL SGGAGTSSVS GPQGDILPVS NSTLKS EDP T GAPVTSSGFL 240
GAPTTLAPGV LEGSEPTAP QPVVTASST PATSPAMVTP QAASQSASGN FMVQVGAVSD 300
QARAQQYQQQ LGQKFGVPGR VTQNGAVWRI QLGP FASKAE ASTLQQRLQT EAQLQS FITT 360
40 AQ 362
<212> Type : PRT
<211> Length : 362
SequenceName : SEQ ID 204
SequenceDescription :

45 Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
50 MKKKTIYQCV ILFFSLLNIH VGMAGPEQVS MHIYGNVVDQ GCDVATKSAL QNIHIGDFNI 60
SDFQAANTVS TAADLNIDIT GCAAGITGAD VLFSGEADTL APTLLKLTDT GGSGGMATGI 120
AVQILDAQSQ QEIPLNQVQP LTPLKAGDNT LKYQLRYKST KAGATGGNAT AVLYFDLVYQ 180
<212> Type : PRT
55 <211> Length : 180
SequenceName : SEQ ID 205
SequenceDescription :

60 Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MKNKLLFMML TILGAPGIAA AAGYDLANSE YNFAVNELSK SSFNQAAIIG QAGTNNSAQL 60
RQGGSKLLAV VAQEGSSNRA KIDQTGDYNL AYIDQAGSAN DASISQ GAYG NTAMITQKGS 120
65 GNKANITQYG TQKTAVVVQR QSQMAIRVTQ R 151
<212> Type : PRT
<211> Length : 151

SequenceName : SEQ ID 206
SequenceDescription :

Sequence

5 <213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MMKFKKCLLP VAMLASFTLA GCQSNADDHA ADVYQTDQLN TKQETKTVNI ISILPAKVAV 60
GGVVVGRVAD ITLDPKTYLP RVTLEIEQRY NHIPDTSSLS IRTSGLLGEQ YLALNVGFED 120
10 EGVSLTYKEG TKVYTSTQEG KECQFTTGLA VVITTYNET RIQPNTKCPE KS 172

<212> Type : PRT

<211> Length : 172

15 SequenceName : SEQ ID 207
SequenceDescription :

Sequence

20 <213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MQTKKNEIIV GIFLLAALLA ALFVCLKAAN VTSIRTESTY TLYATFDNIG GLKARSPVSI 60
GGVVVGRVAD ITLDPKTYLP RVTLEIEQRY NHIPDTSSLS IRTSGLLGEQ YLALNVGFED 120
PELGTAILKD GDTIQDTKSA MVLEDLIGQF LYGSKGDDNK NSGDAPAAAP GNNETTEPVG 180
TTK 183

25 <212> Type : PRT

<211> Length : 183

SequenceName : SEQ ID 208
SequenceDescription :

30 Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MAPLAFSAQS LAESLTVEQR LELLEKALRE TQSELKKYKD EEKKKYTPAT VNRSVSTNDQ 60
35 GYAANPFPTS SAAKPDVAVL KNEEKNASET GSIYSSMTLK DFSKFVKDEI GFSYNGYYRS 120
GWGTASHGSP KSWAIGSLGR FGNEYSGWFD LQLKQRYNE NGKRVDAVVM IDGNVGGQYS 180
TGWFGDNAGG ENFMQFSDMY VTTKGFLPFA PEADFWVGKH GAPKIEIQML DWKTQRTDAA 240
AGVGLENWKV GPGKIDIALV REDVDDYDRS LQNKQQINTH TIDLRYKDIP LWDKATLMVS 300
GRYVTANESA SEKDNQDNNG YYDWKDTWMF GTSLTQKFDK GGFNEFSFLV ANNSIARNFG 360
40 RYAGASPFPT FNGRYYGDHT GGTAVRLTSQ GEAYIGDHFV VANAIVYSFG NNIYSYETGA 420
HSDFESIRAV VRPAYIWDQY NQTGVELGYF TQQNKDANSN KFNESGYKTT LFHTFKVNTS 480
MLTSRLERIF YATYIKALEN ELDGFTFEDN KDAQFAVGAQ AEIWW 525

<212> Type : PRT

<211> Length : 525

45 SequenceName : SEQ ID 209
SequenceDescription :

Sequence

50 <213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MKKRILSAVL VSGVTLSAT TLSAVKADDF DAQIASQDSK INNLTAQQQA AQAQVNTIQG 60
QVSALQTQQA ELQAENQRLE AQSATLGQQI QTLSSKIVAR NESLKQQARS AQKSNAATSY 120
INAIINSKSV SDAINRVSAI REVVSANEKM LQQQEQDKAA VEQKQQENQA AINTVAANQE 180
55 TIAQNTNALN TQQAQLEAAQ LNLQAELETA QDQKATLVAQ KAAAEAAARQ AAAAQAAAEA 240
KAAAEAKALQ EQAAQAQAAA NNNTQATDVS DQQAADANT QAAQTGDSTE QSAAQAVNNS 300
DQESTTATEA QPSASSASTA AVAANTSSAN TYPAGQCTWG VKSLAPWVGN YWNGGGQWAA 360
SAAAAGYRVG STPSAGAVAV WNDGGYGHVA YVTGVQGGQI QVQEANYAGN QSIGNYRGWF 420
NPGSVSYIYP N 431

60 <212> Type : PRT

<211> Length : 431

SequenceName : SEQ ID 210
SequenceDescription :

65 Sequence

<213> OrganismName : Streptococcus mutans UA159

<400> PreSequenceString :
MKVKKTYGFR KSKISKTLTG AVLGTVAAVS VAGQKVFAD E TTTSDVDTK VVGTTQGNPA 60
TNLPEAQSSA SKEAEQSQNQ AGETNGSIPV EVPKTDLDQA AKDAKSAGVN VVQDADVNKG 120
TVKTAEEAVQ KTEIKEDYT KQAEEDIKKT DQYKSDVA AH EAEVAKIKAK NQATKEQY EK 180
5 DMAAHKAEVE RINAANAASK TAYEAKLAQY QADLA AVQKT NAANQAAYQK ALAAYQAE LK 240
RVQEANA AAK AAYDTAV AAN NAKNTE TAA NEIRKRNAT AKA EYETKLA QYQAE LKRVQ 300
EANAANEADY QAKLTAYQTE LARVQKANAD AKAAYEAAVA ANNAKNAALT AENTA IKQRN 360
ENAKATYEAA LKQYEADLAA VKKANAANEA DYQAKLTAYQ TELARVQKAN ADAKAAYEAA 420
VAANNAANAA LTAENTA I KK RNADAKADYE AKLAKYQADL AKYQKDLADY PVKLKAYEDE 480
10 QASIKAAALAE LEKHKNEDGN LTPESAQNLV YDLEPNANLS LT TDGKFLKA SAVDDAFSKS 540
TSKAKYDQKI LQLDDLDITN LEQSN DVASS MELYGNFGDK AGWSTTVSNN SQVKWGSVLL 600
ERGQSATATY TNLQNSYYNG KKISKIVYKY TVDPKSKFQG QKVWLGI FT D PTLGVFASAY 660
TGQVEKNTSI FIKNEFTFYD EDGKPIFN DN ALLSVASLNR ENNSIEMAKD YTGK FVKISG 720
SSIGKNGMI YATDTLNF RQ GQGGARW TMY TRASEPGSGW DSSDAPNSWY GAGAIRMSGP 780
15 NNSVTLGAIS STLVPADPT MAIETGKKPN IWYSLNGKIR AVNVPKVTKE KPTPPVKPTA 840
PTKPTYETEK PLKPAPVAPN YEKEPTPPTR TPQAE PNKP TPPTYETEK P LEPAPVEPSY 900
BAEPTPPTR PDQAE PNKP PPTYETEK EPAPVEPSY AEPTPPTP TP DQPEPNKPE 960
PTYEIPTPP TDPVYQDLPT PPSVPTVHFH YFKLAVQPQV NKEIRNNNDI NIDRTLVAQK 1020
SVVFKQLKTA DLPAGRDETT SFVLVDPLPS GYQFNPEATK AASPGFDVTY DNATNTVTFK 1080
20 ATAATLATFN ADLTKSVATI YPTVVGQVLN DGATYKNNFT LTVNDAYGIK SNVVRVTTFG 1140
KPNDPDNPN NYIKPTKVNK NENGVIDGK TVLAGSTNY Y ELTWDL DQYK NDRSSADTIQ 1200
KGFYYVDDYP EEAL ELRQDL VKITDANGNE VTGVSVDNYT NLEAAPQEIR DVL SKAGIRP 1260
KGAFQIFRAD NPREFYDTYV KTGIDLKIVS PMVVKQMGQ TGGSYENQAY QIDFGNGYAS 1320
NIIINNVPKI NPKKDVT LTL DPADTNNDVG QTIPLNTVFN YRLIGGIIPA DHSEELFEYN 1380
25 FYDDYDQTDG HTGQYKVFA KVDITFKDGS IIKSGAELTQ YTTAEVD TAK GAITIKFKEA 1440
FLRSVSIDSA FQAESYIQMK RIAVGTFENT YINTVNGVTY SSNTVKT TTP EDPTDPTDPQ 1500
DPSSPRTSTV INYKPQSTAY QPSSVQETLP NTGVTNNAYM PLLGIIGLVT SFSLLGLKAK 1560
KD 1562
<212> Type : PRT
30 <211> Length : 1562
SequenceName : SEQ ID 211
SequenceDescription :

Sequence

35 <213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MLTELKAVLK KPMLWITMVG VALVPALYNI IFLSSMWDPY GKVSDLPVAV VNKDKTATYE 60
GKKMTIGKDM TDMVVRNKS L DYHFVDSEKA QKGLEKGDY Y MIITLPEDLS QNAASVLTDE 120
40 PKKLTIPYQT SKGHSFVASK MSETAAKTLK ESVSKNITSS YTKSLFKNMS TLKTGLGSAA 180
NASQKIATGS KQLANGSQVM TDNLNLNSNS SQSFAQGTNT LYSGLTAYTG GVGQLSAGLN 240
NLNNGLTAYT NGVGQLANGS SQLSNQSQKL LGVLAQLANG SASIQQLVNA SSQLNQGLIK 300
LSTATGLSEE VQQQFSSLIN QLGTNLQSIQ NYSDNGTATT ANSPDLSTYL SAITTAQAI 360
VNSGNTSQQT TTNQSNALAA VQATGAYQRL SAEDQSEIAA ALANTGSSTT TTGADANAVS 420
45 QAQAILNNVQ SIQSALSTLQ TTTANTPTSP SASLTQIKNT ANSVLPSAAT SLTTLSSGLT 480
QAKTALDSQV VPVSTALANG TAQLGSTFST GANSIMTG VG QYTNVADILN AGANTLA AKN 540
NQLTDGTSQL VNGANQLNSN SGQLTKGTAQ LANGANQIET GAGKLAAGGE SLTAGLTTLS 600
SGSGELSKAL STAKNKL SLV AVDNDNAKTL SSPVTIKHTD KDNVKTNGVG MAPYMMSAAL 660
MVMAISTNTI FRVALSGKQA KTLREWIDQK LAVNGLIAVT GAILLYFGVH IIGLSANFEL 720
50 KTLGLLIILTS ITFMVLVTTL VTWHDKFGSF AALILLLLQL GSSAGTYPLA VTDKFFQVVN 780
PYLPMSYSVS GLRETISMAG TIGNQLLALS LFFLTFAALG LLIARRRIRS VKVA 834

<212> Type : PRT
<211> Length : 834
55 SequenceName : SEQ ID 212
SequenceDescription :

Sequence

60 <213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MVSQKNKSKK GQSKTFTLIS NRINLLFFLI VALFTVLLLR LAQMQLYDAK FYKSKLTEST 60
TYTIKTSSPR GQIYDAKGVA L VENEVKEVV AFTRSNTMTA KDIKANAKKL ADMVTLTESK 120
VTKRQKDYD LADPKNYQKI VKKLPNKKKY DNFGNNLTES KIYANAVKAV PNSAIDYSED 180
65 EKKI IHIFS MNATSVFN TA SLTTGDLTAE QIAVLATSKS DLKGISVKTD WERKTDKNSI 240
TSIIGKVSSQ KTGLPAEEAN NYVKKGYSLN DRVGTSYLEK QYENDLQGSR TVQAIKVNKE 300
GKIISDKTTA KGTGKGNLKL TLDLEFQKGV EQILNQYFNS ELASGNTKYS EGVYAVVLNP 360

NTGAVLSMAG LEHDLKTGEV SSNALGAVTE VFTPGSVVKG ATLTAGWENG VLSGNQVLND 420
QPIQFAGSSP INSWFTNGST PLTASQSLEY SSNTYMQVLA LKLMGQDYHS GMTLSTDGYK 480
EAMEKLARATY AQYGLGVSTG IDLPGESKGY TPEHYDPSNV LTESFGQFDN YTAMQLAQYA 540
AAVANGGKRI APHLVEGIYD NNKTGGLGNL VQSIDTKVLN NVSISSDDMG I IKEGFYNV 600
5 NGGSYATGKT LAKGASVPIS AKTGTAAYV TGDDGKSVYT SNLNVVAYAP SSNPQIAVAV 660
VLPHTDLHG TTSHAIRDI INLYQKMYPM NQ 692
<212> Type : PRT
<211> Length : 692
SequenceName : SEQ ID 213
10 SequenceDescription :

Sequence

<213> OrganismName : Streptococcus mutans UA159
15 <400> PreSequenceString :
MTVLKYGLGI LLSAILLAI IGGLLFTYYV SSTPKLSEAK LKATNSSLVY DSNNNLIADL 60
GAEKRESISS DSIPMKLVNA VTSIEDHRFF KHRGVDIYRI IGAAWSNLLH KSTQGGSTLD 120
QQLIKLAYFS TKESDQTLKR KAQEVWLSLQ MEKKYTKEEI LTFYVNVKVM GNGNYGMRTA 180
AKSYYGKDLK DLSIAQLATL AGIPQAPTQY DPYAQPKAAT SRRNTVLSQM YKHKKITKRE 240
20 YDAAVATPIS DGLQELKRSS SYPKYMDNYL KQVISEVKKR TGQDIFSAGM KVTYTNVNADA 300
QQYLWNIYNT DEYIAYPDDN FQVASTVMDV TNGKVIAQLG GRHQDTNVSE GTNQAVLTDR 360
DWGSTMKPIS AYGPALSESA FTTTAQMLND SVYYYPGTTT QVYDWDHRYN GWMTIQTAIQ 420
QSRNVPVAVRA IDAAGLD TAK GFLSGLGIDY PEMRYSNAIS SNTSSSEQKY GASSEKMAAA 480
YAAFSNGGT YEPQYVNKIE FKDGTSSETD AKGNRAMKET TAYMMTDMLK TVLTYGTGTE 540
25 AAIPLGLYQAG KTGTSNYDDN ELVEMSEKLG INPYGLGTIA PDENFVGYP QYSMAVWTGY 600
KNRLMPVYGD SMKIAAQVYR TMMAYLSSSG NSDWTMPDGL YRSGGYLYLN GSSGNSRYG 660
AAPATSSSSS SSSSSDSNNN DQNNNQTTTEA SSDSSSSSSD ATTSSNP 707
<212> Type : PRT
<211> Length : 707
30 SequenceName : SEQ ID 214
SequenceDescription :

Sequence

35 <213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MKSKTAKITL LSSSLALAFG ATNVFADEAS TQLNSDTVAA PTADTQASEP AATEKEQSPV 60
VAVVESHTQG NTTTTTSQVT SKELEDAKAN ANQEGLEVTE TEAQKQPSVE AADADNKAQA 120
QTINTAVADY QKAKAEFPQK QEYQNKDFEK YQSDVKEYEA QKAAEQYK EYVAGGLASGR 180
40 VEKAQGLVFI NEPEAKLSIE GVNQYLTKEA RQKHATEDIL QQYNTDNYTA SDFQANPYD 240
PKEDTWFMMK VGDQISVTYD NIVNSKYNDK KISKVKINYT LNSSTNNEGS ALVNLFDHPT 300
KTIFIGAQTS NAGRNDKISV TMQIIFYDEN GNEIDLGN NAIMSLSSLNH WTTKYGDHVE 360
KVNLDGNEFV KIPGSSVDLH GNEIYSAKDN QYKANGATFN GDGADGWDAV NADGTTPRAAT 420
AYYGAGAMTY KGEPFTFTVG GNDQNLPTTI WFATNSAVAV PKDPGAKPTP PEKPELKKPT 480
45 VTWHKNLVE TKTEEVPPVT PPTTPDEPTP EKPKTPEDPQ SPVVAKSVSF RTARKGEMRV 540
RERDYQPTLP HAGAAQNGL ATLGAISTAF AAATLIAARK KEN 583
<212> Type : PRT
<211> Length : 583
50 SequenceName : SEQ ID 215
SequenceDescription :

Sequence

55 <213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MEQKIFSKRK SKIAGLCGAI LTTTVVALAS GTVIEADETI EQPVAAETVS QADGDNPEQT 60
TSVQOETAPQ QTKTSQSSDA TVDSEESATS PSDEQTVSQN DSNSSSQIDQ TIADTNRSDS 120
DHISKTSAAAT TEDQEEKVNS AKAQTAAATN NQDTRYSAKD AYGNSTFNKT LTFEGKNANV 180
ADVTYNGVRD EYIVVNDPSA PYVPNANEIA KYLKEYLTEL RNINNIAIPV PSVDQVMQKY 240
60 AQDRANEEN EKNGLDHDN LPIPNLTLTW AEDGHLDMDS SIQSKSQEGY TLASDKATAY 300
YLALNWFSDY FNIYDDPNDG LKSFHGAIVI LSDGGTGMGL GLASGQDNEK GMWYAQLEFG 360
GNDNEDNTND FSSSLKNGKGE WVLYYKGSVP KFLPNTTFWY VKKGTSPDAA STPHNSDKPS 420
FQSSKDLDPN FKADNRFQEG KEASVHQAIP ATFKSHRDEV GNKDQNSLSA QLPDGTGVQKN 480
NQLALIALGT GLILLSGLLL SKRKSLK 507
65 <212> Type : PRT
<211> Length : 507
SequenceName : SEQ ID 216

SequenceDescription :

Sequence

5 <213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MTFEKQKHFS LRKLKFLVLS VAIIAFLFAV TKTAEADETV ITEQRQTSKI NASSQKVENQ 60
TSNQVEAKTD SANKDPQEK T GSVATDAPSM NSANNMSQSD KQNTVNEISS DSQQTKTDEQ 120
TDLFPQNSFKQ QSAHVKMTTE AEKTPSHSIN TFFVNDGNGNW YYLGADGRNV TGSHTIGGKT 180
10 MYFAQDGKQV KGAFADSDG NKHYDRDSG EMWTRNFVND QGNWYYLNND GVPVGTGSITV 240
NGQSLYFNDS GSQVKGNFVE EDGSLRYDK NSGDLRLKTS RTINGVNYQF DNDGNARAID 300
KIEVVKTSLV VDSYEFGPSV SKIILEFNHK VTPAVVHAGA MVTTAGVQRK ILNSYVSNAS 360
GHVVYFDSSH YVLELDIPY DPNDSSRNAS PFIFDSAAFR NNWVNSYTVK VDNLQVQADG 420
SNSSQIISSE QDAINNRFPL TTDRFSEGRS YGNFNYYAAYQ PEAAIGGEKN PLIVWLHGIG 480
15 EVGTDINIP LASNVARL TE DPIQSHFTST GSGGQKGAYV LVPQSSIPWS QNQTASLMAL 540
IKAYVASHPD IDSRRYLAG VSNGGGMTLD MGVAYPNYFA ALVPPIAASYS NQLTDNQITA 600
AALKALKGQP MWLIHTRTDK TISADSSVLP FYKELLQAGA QNKWLSYET NVGKHHSQVT 660
YNGHWSWIYF LNDQVTGTQN TDNAKNWSGL SGMVATNPTY GGDATATVNG RTYSNVFDWL 720
NGQRRR 726
20 <212> Type : PRT
<211> Length : 726
SequenceName : SEQ ID 217
SequenceDescription :

Sequence

25 <213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MKIFIKKHQQ SILYYSLSFL LPSFIMFLVL FSKNIYWGSS TTILASDGPH QYVIFDALFR 60
30 NILHGTDSL FYSFAGLGFN IFALTSYYLG SFLTPPTYFF NVKNMADAFY LFTLIKFGLI 120
GLSAFYSLGQ IYTKISKSLV LMLSTSYALM SFTSSQLELN NWLDVFI LLP LIMLGLQRLV 180
EKRGIFLYFL TLTCFLIQNY YFGFMTAIFL TLWFFTQVSW DIRNRMKRLS DFVLVSIFAT 240
LTSAFMLLPT FLDLKSHEV LTEQISLFSS DIWYDFFAK SLLGSYDTTK YGSIPTIYIG 300
35 LLPLIFAITE FVKSIIKQV KVAYFLLLAI IASFIQPL DLFWQGMHSP NMFLHRYSWA 360
FSLVIVIMAA ETLTRIKDIK LKNFYPAFTF LGVGLLATFL FKDYNNYLTQ VNFILTTIFL 420
VSYFILLTFE FNQLVSYKVI ISFTLIFTSE EIALNTFYQI EGIQTDWNEP SREVYEDNVK 480
EIDNVVKKTK KDNLEFFRTE KQIPQTYNDG MKFNYSISQ FSSVKNNLSA QLLNSLGYYS 540
QGNHSTISYP NNTILMDSL FSIKYNINQN PHKFGFHLKQ KNNKLQLYKN FYSPLPLALMS 600
NHIYKDVKFD SYPLDNQKQF VNELTDLNLT LFKBIPITSS VGMQVLDNRV TINGSKGNKA 660
40 QVYTVKCPA NSQLYISLPN LTVNNKDNV FITNKHSTSS YIIDESYYLF NLGNYKKTQT 720
LIFKLSPKN KTVSYDLPHI YALDLTAYQK SIKQLKSQTV KTTTKKNKIF TTYVAKKRTS 780
LIYTLPPYDKG WFAKQNGKAI KISKAQNGLM KIDVSKGSGK IIMTFVPQGL YQGILLTCLG 840
IFLFVFPQLY YKKFNLK 857
45 <212> Type : PRT
<211> Length : 857
SequenceName : SEQ ID 218
SequenceDescription :

Sequence

50 <213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MKLKHILRIG AVAFASILL TACGSKTSKK TVTLATVGTT NPFSYEKKGK LTGYDIEVAK 60
EVFKASDKYD VKYQKTEWTS IFSGLDSDKY QIGANNISYT KERANKYLYS NPTASNPLVL 120
55 VVPKDSDIKS YNDIAGHSTQ VVQGNITVSM LQKFNKNHEN NQVKLNFTSE DLAHQIRNVS 180
DGKYDFKIFE KISAETIIE QGLDNLKVID LPSDQKPYVY FIFAQDQKDL QKFVNKRLKK 240
LYENGTLEKL SKKYLGGSYL PDKKDMK 267
60 <212> Type : PRT
<211> Length : 267
SequenceName : SEQ ID 219
SequenceDescription :

Sequence

65 <213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MRFLVFLIAF FAFYKFIET ERIDSNVAV NPDSLILKRF LKTNQLNGIM IVTGPDKAQ 60

VFSNQSKVDG SPVSIKDYFP LASLQKLITG VAIQQLIDKG KLSLNTPLSK YYPQIENSEN 120
ITIQNLLTHT SGLADRKEVP QQVLTTQEQQ LDFSLTNYRV TYRKKWKYAN INYALLAGII 180
SQISGQNYAT YVRQHFLTAG KGWHFKKYYIQ IKDKSKLAAL SVMQDSTTWD KLSKEVTSTF 240
GAGDYASRPV DYWKFMMAFI NDQFVPVSEY QRSMKMTSKS YYGGLYISQK MLHANGGGFD 300
5 TYSCFAYSNP KTKQVMVLF I TNGKYKRVKS LAAKAFKLYA DSYALRKNET SK 352

<212> Type : PRT
<211> Length : 352
SequenceName : SEQ ID 220
10 SequenceDescription :

Sequence

<213> OrganismName : Streptococcus mutans UA159
15 <400> PreSequenceString :
MKKKIALAAL SFVSAAVLAA CSSAPGGSSD AAGNKIGDTV KIGYNLELSG DVAAYGQAEK 60
NGANLAVEEI NKAGGIDGKK IKVISKDNKS DNGEASTIST NLATQSKVNA ILGPATSGAT 120
AAAAPNANDA AVPLIVTPSGT QDNLTYSKKG VQDYIFRTTF QDSFQGGKIIA KYATDNLKAK 180
KVALYYDKSS DYAQGIADAF KKAYKGKITV EDTFQAKDQD FQAALTKFKN KDFDAIVIPG 240
20 YYTETGLITK QARDMGLTQP ILGPDGFNDE KYVEGAGAAN TNNVHYVSGY STKVALTNKA 300
BKFLKDYKAK YGEEPNMFAA LAYDSVYMLA DAAKDAKTSK DIATNLAKLK NFKGVTGKMT 360
IDKKHNPVKS AVMVGLKDGK EDTATAVEAK 390
<212> Type : PRT
<211> Length : 390
25 SequenceName : SEQ ID 221
SequenceDescription :

Sequence

<213> OrganismName : Streptococcus mutans UA159
30 <400> PreSequenceString :
MKKLSLLLLV CLSLLGLFAC TSKKTADKKL TVVATNSIIA DITKNIAGNK VVLHSIVPVG 60
RDPHEYEPLP EDVKKTSQAD VIFYNGINLE NGGNAWFTKL VKNNAHKKTDK DYFAVSDSVK 120
TIYLENAREK GKEDPHAWLD LKNGIYAKN IMKRLSEKDP KNKSYYQKNF QAYSAKLEKL 180
35 HKVAKEKISR IPTCKMIVT SEGCFKYFSK AYDIPSAYIW EINTEEEGTP NQIKALVKKL 240
RKSRVSALEFV ESSVDDRPMK TVSKDTGIPI AAKIFTDSVA KKGQAGDSYY AMMKWNIDKI 300
ANGLSQ 306
<212> Type : PRT
<211> Length : 306
40 SequenceName : SEQ ID 222
SequenceDescription :

Sequence

<213> OrganismName : Streptococcus mutans UA159
45 <400> PreSequenceString :
MFVHTTKTKKK RKWQRKVFL LLLFLLPIVS VLAFIGVLFVGGTAESHVVE ATTGGVKLSA 60
KQFADKTKLG ISEEEAKNAL AFADRLMSRH HFTAQATAGV LAVGFRESGF DVKAVNNSGG 120
VAGFFQWGSW GSSVNGDRWK VASKRELTL EVDLMSTEL DGRYADVVKK VGSATDEKQA 180
50 AKDWSQYYEG VAVSDGQTKA DKIESWATTI CEALKSGGTN YAKVNNTGTS STAIPQGWEN 240
ISAFDGHAYE GSENVPOGQC TWYVYNRAKQ LGVSFSPYMG NGGQWYQVQG YHSSHTPKAH 300
TALS FVNGQA GSDPTYGHVA FVEAVKDDGS ILISEMNVYG QPAMTVAYRT FDAETAKQFW 360
YVEGK 365
<212> Type : PRT
55 <211> Length : 365
SequenceName : SEQ ID 223
SequenceDescription :

Sequence

<213> OrganismName : Streptococcus mutans UA159
60 <400> PreSequenceString :
MKMKRKL LSL VSVLTILLGA FWVTIKIVKAD QVTNYTNTAS ITKSDGTALS NDPSKAVNYW 60
EPLSFSNSIT FPDEVSIGAG DTLTIKLPEQ LQFTTALTFD VMHTNGQLAG KATTDPN'TGE 120
65 VTVTFTDIFE KLPNDKAMTL NFNAQLNHN ISIPGVVNFN YNNVAYSSV KKDITPISP 180
DVNKVGYQDK SNPGLIHWKV LINNKQGAID NLTLTDVUGE DQETVKDSL V AARLQYIAGD 240
DVDSLDEAAS RPYAEDFSKN VTYQTNDLGL TTGFTYTIPG SSNNAIFISY TTRLTSSQSA 300

GKDVSNITIAI SGNNINYSNQ TGYARIESAY GRASSRVKRQ AETTTVTETT TSSSETTTS 360
EATTETSSST NNNSTTTETA TSTTGASTTQ TKTTASQTNV PTTTNIITTS KQVTKQKAKF 420
VLPSTGEQAG LLLTTTVGLVI VAVAGVYFYR TRR 453
<212> Type : PRT
5 <211> Length : 453
SequenceName : SEQ ID 224
SequenceDescription :

Sequence

10 <213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MTFKKLVLGL LSFVAVFTLV ACSSSNSKNL QDDIKEKKKL VVAVSPDYAP FEFKALVNGK 60
DTVVGADIDL AKAIKAKELGV KLELSSMSFD NVLSSLKTGK ADIAISGLSY TKERAQAYDF 120
15 SEAYYKTENA ILIKKSDLNK YTMISFNNK TKVAVQKGTI EEGLAKNQLK QSNITSLTSM 180
GEAVNELKSG QVDAIDLEKP VAEGYVSQNS DLVLAKVALK TEGGDAKAVA LPKDSGQLVK 240
TVNKVIKKLK KEDKYKQFIS DAVKLTGQQV D 271
<212> Type : PRT
<211> Length : 271
20 SequenceName : SEQ ID 225
SequenceDescription :

Sequence

25 <213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MKKHFFMTFS LLLAAVFLVA CSNLSDSGQR NWDKINKRCM LKIATAGTLY PQSYHDDHNNK 60
LTGYDVEILK EIGKRLGLKV QFTEMGVDGM LTAIKSGQID VANYSLEDGN KNISKFLRTS 120
PYKYSFTSMV VRSKDDSGIH SWSDLKGGKA AGAASNMYK IAKKLGAKLV VYDNVTNDVY 180
30 MKDLVNGRTD VIINDYYLQK IAVAAVKDKY AIKINQGLYA NPYSTSFTLS LKNKVLQKKI 240
NKAVKDMRKD GTLTKLSKKF FQGEDVTKKH YNSYKKIDIS DVD 283
<212> Type : PRT
<211> Length : 283
SequenceName : SEQ ID 226
35 SequenceDescription :

Sequence

40 <213> OrganismName : Streptococcus pneumoniae R6
<400> PreSequenceString :
MKLLKKMMQV ALATFFFGLL GTSTVFADDS EGWQFVQENG RTYYKKGALK ETYWRVIDGK 60
YYFDPPLSGE MVVGWQYIPA PHKGVITGPS PRIEIALRPD WFYFGQDGLV QEFVGKQVLE 120
AKTATNTNKH HGEEYDSQAE KRVYFEDQR SYHTLKTGWI YEEGYWYYLQ KDGGFDSRIN 180
RLTVGELARG WKDYPLTYD BEKLKAAPWY YLDPATGWQN LGNKWYYLRS SGAMATGWYQ 240
45 EGSTWYYLNA SNGDMKTGW FQVNGNWYYAY DSGALAVNTT VGGYYLNYNG EWWK 294
<212> Type : PRT
<211> Length : 294
SequenceName : SEQ ID 227
50 SequenceDescription :

Sequence

55 <213> OrganismName : Streptococcus pneumoniae R6
<400> PreSequenceString :
MKLLKKMMQV LLAVFFFGLL ATNTVFANTT GGRFVDKDNR KYYVKDDHKA IYWHKIDGKT 60
YYFGDIGEMV VGWQYLEIPG TGYRDNLFDN QPVNEIGLQE KWWWFGQDGA LLEQTDKQVL 120
EAKTSENTGK VYGEQYPLSA EKRTYFDDNN YAVKTGWIYE DGNWYYLNKL GNFGDDSYNP 180
LPIGEVAKGW TQDFHVTIDI DRSKPAPWYY LDASGKMLTD WQKVNGKWYY FGSSGSMATG 240
60 WKYVRGKWWY LDNKNKGMKT GWQYLGKWKY YLRSSGAMVT GWYQDGLTWY YLNAGNGDMK 300
TGWFQVNGKW YYAYSSGALA VNTTVDGYSV NYNGEWWQ 338
<212> Type : PRT
<211> Length : 338
SequenceName : SEQ ID 228
65 SequenceDescription :

Sequence


```
-----
<213> OrganismName : Streptococcus pneumoniae R6
<400> PreSequenceString :
5 MNKKKMILTS LASVAILGAG FVASQPTVVR AEESPVASQS KAEKDYDAAK KDAKNAKKAV 60
  EDAQKALDDA KAAQKKYDED QKKTEEKAAL EKAASEEMDK AVAAVQQAYL AYQQATDKAA 120
  KDAADKMIDE AKKREEEAKT KFNTVVRAMVV PEPEQLAETK KKSEEAQKKA PELTKKLEEA 180
  KAKLEEAEEK ATEAKQKVDA EEVAPQAKIA ELENQVHRLE QELKEIDESE SEDYAKEGFR 240
  APLQSKLDAK KAKLSKLEEL SDKIDELDAE IAKLEDQLKA AEENNNVEDY FKEGLEKTIA 300
  AKKAELEKTE ADLKKAVNEP EKPAPAPETP APEAPAEQPK PAPAPQPAPA PKPEKPAPQP 360
10 KPEKTDQQA EEDYARRSEE EYNRLTQQQP PKAEKPAPAP KTGWKQENGM WYFYNTDGS 420
  ATGWLQNNGS WYYLNSNGAM ATGWLQYNGS WYYLNANGAM ATGWAKVNGS WYYLNANGAM 480
  ATGWLQYNGS WYYLNANGAM ATGWAKVNGS WYYLNANGAM ATGWLQYNGS WYYLNANGAM 540
  ATGWAKVNGS WYYLNANGAM ATGWVKDGDW WYYLEASGAM KASQWFKVSD KWYYVNGLGA 600
  LAVNTTVDGY KVNANGEWV 619
15 <212> Type : PRT
  <211> Length : 619
      SequenceName : SEQ ID 229
      SequenceDescription :

20 Sequence
  -----
  <213> OrganismName : Streptococcus pneumoniae R6
  <400> PreSequenceString :
25 MKILPFIARG TSYLLKMSVK KLVPLVVGVL MLAAGDSVYA YSRGNGSIAR GDDYPAYYKN 60
  GSQEIDQWRM YSRQCTSFVA FRLSNVNGFE IPAAYGNANE WGHRRAREGY RVDNPTTIGS 120
  ITWSTAGTYG HVAWVSNVMG DQIEIEEYNY GYTESYNKRIV IKANTMTGFI HFKDLDSGSV 180
  GNSQSSASTG GTHYFKTKSA IKTEPLVSAT VIDYYPGEEK VHYDQILEKD GYKWLSTAY 240
  NGSYRYVQLE AVNKNPLGNS VLSSTGGTHY FKIKSAIKTE PLVSATVIDY YYPGEKVHYD 300
  QILEKDGKWK LSYTAYNGSR RYIQLEGVTS SQNYQNQSGN ISSYGSNNSS TVGWKKINGS 360
30 WYHFKSNGSK STGWLKDGSS WYYLKLSEGM QTGWLKENG S WYYLGSSGAM KTGWYQVSGE 420
  WYYSYSSGAL AINTTVDGYR VNSDGERV 448
  <212> Type : PRT
  <211> Length : 448
      SequenceName : SEQ ID 230
      SequenceDescription :

35 Sequence
  -----
  <213> OrganismName : Streptococcus pneumoniae R6
  <400> PreSequenceString :
40 MFASKSERKV HYSIRKFSIG VASVAVASLV MGSVVHATEN EGSTQAATSS NMAKTEHRKA 60
  AKQVVDEYIE KMLREIQLDR RKHTQNVALN IKLSAIKTKY LRELNVLEEK SKDELPSIEK 120
  AKLDAAEFKF KKDITLKPGEK VAEAKKKVEE AKKKAEDQKE EDRRNYPNT YKTLELEIAE 180
  FDKVKKEAEEL ELVKBEAKES RNEGTLKQAK EKVESKKAEA TRLENIKTDR KKAEEEAARK 240
45 ADAKLKEANV ATSDQGKPKG RAKRGVPGEL ATPDKKENDA KSSDSSVGEE TLPSSSLKSG 300
  KKVAAEAEKKV EEAEEKAKDQ KEEDRRNYPT NTYKTLGLEI AESDVKVKEA ELELVKEEAK 360
  EPRDEEKIKQ AKAKVESKKA EATRLENIKT DRKKAEEEA RKAEEEDKVK EKPAEQPPA 420
  PATQPEKPAP KPEKPAPQPK AEKTDQQA EEDYARRSEEE YNRLTQQQPP KTEKPAQPST 480
  PKTGWKQENG MWYFYNTDGS MATGWLQNNG SWYYLNANGA MATGWLQNNG SWYYLNANGS 540
50 MATGWLQNNG SWYYLNANGA MATGWLQYNG SWYYLNSNGA MATGWLQYNG SWYYLNANGD 600
  MATGWLQNNG SWYYLNANGD MATGWLQYNG SWYYLNANGD MATGWVKDGD TWYYLEASGA 660
  MKASQWFKVS DKWYYVNGSG ALAVNTTVDG YGVNANGEWV N 701
  <212> Type : PRT
  <211> Length : 701
      SequenceName : SEQ ID 231
      SequenceDescription :

60 Sequence
  -----
  <213> OrganismName : Streptococcus pneumoniae R6
  <400> PreSequenceString :
  MKKTTILSLT TAAVILAAYV PNEPILADTP SSEVIKETKV GSIIQQNNIK 60
  YKVLTVENGI GTVQVGVNGVT PVEFEAGQDG KPFTIPTKIT VGDVFTVTE VASQAFSYYP 120
  DETGRIVYYP SSTIPSSIK KIQKKGFHGS KAKTIIFDKG SQLEKIEDRA FDFSELEIE 180
65 LPASLEYIGT SAFSFSQKLE KLTFFSSSKL ELISHEAFAN LSNLEKLTLP KSVKTLGSNL 240
  FRLTSLKHV DVEEGNESFA SVDGVLFSD KTQLIYPSQ KNDESYPKPK ETKELASYSF 300
  NKNSYLKLE LNEGLEKIGT FAFADAIKLE EISLPNSLET IERLAFYGNL ELKELILPDN 360
```

VKNFGKHVMN GLPKFLTLTG NNINSLPSFF LSGVLDLKE IHIKNKSTEF SVKKDTFAIP 420
ETVKFYVTSE HIKDVLKSNL STSNDIIVEK VDNIKQETDV AKPKKNSNQ VVGWVKDKGL 480
WYYLNEGSM ATGWVKDKGL WYYLNEGSM ATGWVKDKGL WYYLNEGSM ATGWVKDKGL 540
WYYLNEGSM ATGWVKDKGL WYYLNEGSM ATGWVKDKGL WYYLNEGSM ATGWVKDKGL 600
5 WYYLNEGSM ATGWVKDKGL WYYLNEGSM ATGWVKDKGL WYYLNEGSM ATGWVKVSGK 660
WYYTYNSGDL LVNTTTPDGY RVNANGWVG 690
<212> Type : PRT
<211> Length : 690
SequenceName : SEQ ID 232
10 SequenceDescription :

Sequence

<213> OrganismName : Streptococcus pneumoniae R6
15 <400> PreSequenceString :
MEINVSKLRT DLPQGVQPY RQVHAHSTGN PHSTVQNEAD YHWRKDPELG FFSHIVGNCG 60
IMQVGPVDNG AWDVGGGUNA ETYAAVELIE SHSTKEEFMT DYRLYIELLR NLADEAGLPK 120
TLDTGSLAGI KTHEYCTNNQ PNNHSDHVDY YPYLAKWGIS REQFKHDIEN GLTIETGWQK 180
NDTGYWYVHS DGSYPKDKFE KINGTWYFYD SSGYMLADRW RKHTDGNWYW FDNSGEMATG 240
20 WKKIADKWY FNEEGAMKTG WVKYKDTWY LDAKEGAMVS NAFIQSADGT GWYYLKP DGT 300
LADRPEFTVE PDGLITVK 318
<212> Type : PRT
<211> Length : 318
SequenceName : SEQ ID 233
25 SequenceDescription :

Sequence

<213> OrganismName : Neisseria meningitidis Z2491
30 <400> PreSequenceString :
MTFAYWCILI AYLLPLFCAA YAKKAGGFRF KDNHNPRDFL ARTQGTAAARA HAAQONGFEA 60
FAPFAAVLT AHATGNAGQA TVNTLAGLFI LERLAFIWCY IADKAALRSL MWVGGFVCTV 120
GLFVVAA 127
<212> Type : PRT
35 <211> Length : 127
SequenceName : SEQ ID 234
SequenceDescription :

Sequence

40 <213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MNKIYRIWN SALNAWVAVS ELTRNHTKRA SATVKTAVLA TLLFATVQAN ATDEDEEEEL 60
ESVQRSVVG IASMEGSGE LETISLSMTN DSKEFVDPYI VVTLKAGDNL KIKQNTNENT 120
45 NASSFTYSLK KDLTGLINVE TEKLSFGANG KKVNIISDTK GLNFAKETAG TNGDTTVHLN 180
GIGSTLDTL AGSSASHVDA GNQSTHYTRA ASIKDVLNAG WNIKGVKTGS TTGQSENVD 240
VRTYDTVEFL SADTKTTTVN VESKDNGKRT EVKIGAKTSV IKEKDGLVT GKKGENGSS 300
TDEGEGLVTA KEVIDAVNKA GWRMKTITAN GQTGQADKFE TVTSGTNVTF ASGKGTTATV 360
SKDDQGNITV MYDVNVGDAL NVNQLQNSGW NLDSKAVAGS SGKVISGNVS PSKGKMDTV 420
50 NINAGNNIEI SRNGKNIDIA TSMAPQFSSV SLGAGADAPT LSVDDGALN VGSKDANKPV 480
RITNVAPGVK EGDVTNVAQL KGVAQNLNNR IDNVDGNARA GIAQAIATAG LVQAYLPKGS 540
MMAITGGGTYR GEAGYAIGYS SISDGGNWII KGTASGNSRG HFGASASVGY QW 592
<212> Type : PRT
55 <211> Length : 592
SequenceName : SEQ ID 235
SequenceDescription :

Sequence

60 <213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MLLAEGQKSA VTEYYLNHGT WPSNNSDAGV ASTATDIKKG YVKEVKVEKG VITATMLSSG 60
VNNEIKGKKL SLWAKRQAGS VKWFCGQPV E RAANNAANDA VTAATANGNG KIDTKHLPST 120
65 CRDAASAVCI ETPPTAFYKN T 141
<212> Type : PRT
<211> Length : 141

SequenceName : SEQ ID 236
SequenceDescription :

5 Sequence

<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MKTDDKRTTE THRKAPKTGR IRFSPAYLAI CLSFGILPQA WAGHTYFGIN YQYYRDFAEEN 60
KGKFAVGAKE IEVYNKKGEL VGKSMTKAPM IDFSVVSRRNG VAALVGDQYI VSVAHNGGYN 120
10 NVDFGAEGRN PDQHRFSYQI VKRNNYKPDN SHPYNGDYHM PRLHKFVTDA EPVEMTSDMR 180
GNTYSDKEKY PERVRIGSGH HYWRYDDDKH GDLSYSGAWL IGGNTHMQGW GNNGVVSLSG 240
DVRHANDYGP MPIAGAAGDS GSPMFTYDKT NNKWLNLGVL QTGYPYSGRE NGFQLIRKDW 300
FYDDIYRGDT HTVFFEPFRSN GHFSFTSNNN GTGTVTETNE KVSNNPKLVQ TVRLFDESIN 360
ETDKPEPVYAA GGVNQYRPRL NNGENLSFID YGNGKLILSN NINQGAGGLY FEGDFTVSPE 420
15 NNETWQAGAV HISEDSTVTW KVNQVANDRL SKIGKGTLLHV QAKGENQSSI SVGDGTIVILD 480
QQADDKGKKQ AFSEIGLVSG RGTVQLNADN QFNPDKLYFG FRGGRLDLNG HSLSFHRIQN 540
TDEGAMIVNH NATTTSTVTI TGNESITQPS GKNINRLNYS KEIAYNGWFG EKDTTKTNGR 600
LNLVYQPAAE DRTLLLSGGT NLNGNITQTN GKLEFFSGRPT PHAYNHLGSG WSKMEGIPQG 660
EIVWDNDWIN RTFKAENFHI QGGQAVISRN VAKVEGDWHL SNHAQAVFGV APHQSHITCT 720
20 RSDWTGLTNC VEKTTTDDKV IASLTKTDIS GNVSLADHAH LNLTLGLATLN GNLSANGDTR 780
YTVSHNATQN GNLSLVGNAQ ATFNQATLNG NTSASGNASF NLSNNAQNG SLTSLDNAKA 840
NVSHSALNGN VSLADKAVFH FENSRTFQQL SGSKDTALHL KDSEWTLPSG TELGNLNLNDN 900
ATITLNSAYR HDAAGAQTGS VSDTPRRRSR RSLSVTPPT SVESRENTLT VNGKLNQGT 960
FRFMSLELFGY RSDKLKLAES SEGTYTLAVN NTGNEPVSLE QLTVEGKDN KPLSENLENT 1020
25 LQNEHVDAGA WRYQLIRKDG EFRLHNPVKE QELSDKLGA EAKKQAEKDN AQLSDALIAA 1080
GRDAAEKTES VAEPARQAGG ENVGIMQAE EKKRVQADKD SALAKQREAE TRPATTAFPR 1140
ARRARDLPQ PQPQPQPQPQ PQRDLSRYA NSGLSEFSAT LNSVFAVQDE LDRVFAEDRR 1200
NAVWTSIGIRD TKHYRSQDFR AYRQQTDLRQ IGMQKNLGS RVGILFSHNR TENTFDDGIG 1260
NSARLAHGAV FGQYIGIRFD IGISTGAGFS SGLSDGIGG KIRRRVLHYG IQARYRAGFG 1320
30 GFGIEPYIGA TRYFVQKADY RYENVNIATP GLAFNRYRAG IKADYSFKPA QHISITPYLS 1380
LSYTDAASGK VTRTVNTAVL AQDFGKTRSA EWGVNAEIKG FTLSLHAAAA KGPQLEAQHS 1440
AGIKLGYRW 1449

<212> Type : PRT

<211> Length : 1449

35 SequenceName : SEQ ID 237
SequenceDescription :

Sequence

40 -----
<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MNTLQKGFTL IELMIVIAIV GILAAVALPA YQDYTARAQV SEAILLAEGQ KSAVTEYYLN 60
HGEWPSNNTS AGVASSTDIK GKYVQSVVEK NGVVTATMAS SNVNNEIKGK KLSLWAKRQD 120
GSVKWFCCQP VKRNDTATTN DDVKADTAAN GKQIDTKHLP STCRDAASAG 170
45 <212> Type : PRT
<211> Length : 170
SequenceName : SEQ ID 238
SequenceDescription :

50 Sequence

<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MQARLLIPIL FSVFILLSACG TLTGIPSHGG GKRFAVEQEL VAASARAQVK DMDLQALHGR 60
55 KVALYIATMG DQSGSGLTGG RYSIDALIRG EYINSPAVRT DYTYPREYET AETTSGLTGT 120
LTTSLSTLNA PALSRQSDG SGSKSSLGLN IGGMGDYRNE TLTNPRDTA FLHLVQTVF 180
FLRGIDVVSP ANADTDVFIN IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL 240
IKPKTNAFEA AYKENYALWM GPYKVSIGIK PTEGLMVDFA DIQPYGNHMG NSAPSVEADN 300
SHEGYGSDE AVRRHRQGP 320
60 <212> Type : PRT
<211> Length : 320
SequenceName : SEQ ID 239
SequenceDescription :

65 Sequence

<213> OrganismName : Neisseria meningitidis Z2491

<400> PreSequenceString :
MRPIFLSFVL FPILITACST PDKSARWENI GTISNGNIHT YINKDSVRKN GNLMIQDKK 60
VVTNLKQERF ANTPAYKTAI AEWEIHCNNK TYRLSSLQLF DTKNTEISTQ NYTASSLRPM 120
SILSGTLTEK QYETVCGKKL 140

5 <212> Type : PRT
<211> Length : 140
SequenceName : SEQ ID 240
SequenceDescription :

10 Sequence

<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MNKLFITALS ALALSACAGT WEGAKQDTAR NLDKTQAAAE RAAEQTGNV EKGWDKTKEA 60
15 VKKGGNAVGR GISHLGGKIE NATE 84
<212> Type : PRT
<211> Length : 84
SequenceName : SEQ ID 241
SequenceDescription :

20 Sequence

<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
25 MKLLFIPLVL FVAVEHFYIA WLEMTQIPSE KAAETFKLPY EFMEQNRVQT LFGNQGLYNG 60
FLGIGLVWSR FAAPDNAVYG ATVLFLGFVL IAAAWGAFSS GNKGILVKQG LPAFLAAAV 120
LAV 123
<212> Type : PRT
<211> Length : 123
30 SequenceName : SEQ ID 242
SequenceDescription :

Sequence

35 <213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MASSNVNNEI KDKKLSLWAK RQDGSVKWFC GQPVKRDAAT DADVTADSGN EIDTKHLPST 60
CRDAASAVCT KTPYEYPNHG EWPKNFVIPA QAGIQVCRHG NLSGKKVSPV LSSRFPLSWE 120

40 <212> Type : PRT
<211> Length : 120
SequenceName : SEQ ID 243
SequenceDescription :

45 Sequence

<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MLLAEGQKSA VTEYYLNHGE WPSNNTSAGV ATSTDIKGY VQSVEVKNGV VTATMASSNV 60
50 NNEIKGKKLS LWAKRQDGSV KWFCGQPVKR NDTATTNDDV KADTAANGKQ IDTKHLPSTA 120
STRKSTPN 128
<212> Type : PRT
<211> Length : 128
SequenceName : SEQ ID 244
55 SequenceDescription :

Sequence

60 <213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MPIPFKPVLA AAAlAQAFPA FAADPAPQSA QTLNEITVTG THKTQKLGE E KIRRKTLCLK 60
LVNDEHDLVR YDPGISVVEG GRAGSNGFTI RGVDKDRVAI NVDGLAQAES RSSEAFQELF 120
GAYGNFNANR NTSEPFNFSE VTITKGADSL KSGSGALGGA VNYQTKSASD YVSEDKPYHL 180
GIKGGSVGKN SQKFSSITAA GRLFGLDALL VYTRRFGKET KNRSTEGDIE IKNDGYVYNP 240
65 TDTGGPSKYL TYVATGVARS QPDPQEWVKN STLFLKGYNF NDQNRIGWIF EDSRTDRFTN 300
ELSNLWTGTT TSAATGDYRH RQDVSYYRRS GVEYKNELEH GPWDSLKLRY DKQRIDMNTW 360
TWDIPKNYDK RGINEGEVYHS FRHIRQNTAQ WTADFEKQLD FSKAVWAAQY GLGGGKGDNA 420

```

NSDYSYFAKL YDPKILASNQ AKITMLIENR SKYKFAYWNN AFHLGGNDRF RLNAGIRYDK 480
NSSSAKDDPK YTTAIRGQIP HLGSERAHAG FSYGTGFDWR FTKHLHLLAK YSTGFRAPTS 540
DETWLLFPHP DFYLKANPNL KAEKAKNWEL GLAGSGKAGN FKLSGFKTKY RDFIELTYMG 600
VSSDDKNNPR YAPLSDGTAL VSSPVWQONQ RSAAWVKGIE FNGTWNLDSI GLPKGLHTGL 660
5 NVSYIKGKAT QNNGKETPIN ALSPWTAVYS LGYDAPSKRW GINAYATRTA AKKPSDTVHS 720
NDDLNNPWPY AKHSKAYTLF DLSAYLNIGK QVTLRAAAYN ITNKQYYTWE SLRSIREFGT 780
VNRVDNKTHA GIQRFTSPGR SYNFTIEAKF 810
<212> Type : PRT
<211> Length : 810
10 SequenceName : SEQ ID 245
SequenceDescription :

Sequence
-----
15 <213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MKKSLLIALL AALPVAAMAD VTLYGTIKTG VETSRSEVEHN GGQVSVVETG TGIVDLGSKI 60
GFKGQEDLGN GLKAIWQVEQ KASIAGTDSG WGNRQSFGL KGGFGKLRVG RLNSVLKDTG 120
DINPWDSKSD YLGVNKIAEP EARLISVRYD SPEFAGLSGS VQYALNDNVG RHNSSESYHAG 180
20 FNYKNGGFFV QYGGAYKRHQ DVDDVKIEKY QIHRLVSGYD NDALYASVAV QQQDAKLVED 240
NSHNSQTEVA ATLAYRFGNV TPRVSYAHGF KGSVDDAKRD NTYDQVVVGA EYDFSKRTSA 300
LVSAGWLQEG KGENKFVATA GGVGLRHKF 329
<212> Type : PRT
<211> Length : 329
25 SequenceName : SEQ ID 246
SequenceDescription :

Sequence
-----
30 <213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MKTLLLLLIPL VLTACGTLTG IPAHHGGKRF AVEQELVAAS SRAAVKEMDL SALKGRKAAL 60
YVSVMGDQGS GNISGGGRYSI DALIRGGYHN NPESATQYSY PAYDTTATTK SDALSSVTTS 120
TSLLNAPAAA LTKNSGRKGE RSAGLSVNGT GDYRNETLLA NPRDVSFLTIN LIQTVFYLRG 180
35 IEVVPPEYAD TDVVFVTVDF GTVRSRTELH LYNAETLKAQ TKLEYFAVDR DSRKLLIAPK 240
TAAYESQYQE QYALWMPYS VGKTVKASDR LMVDFSDITP YGDTTAQNRP DFKQNGKKP 300
DVGNEVIRRR KGG 313
<212> Type : PRT
<211> Length : 313
40 SequenceName : SEQ ID 247
SequenceDescription :

Sequence
-----
45 <213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MNKTLILPV AILLGGCAAG GGNTFGSLDG GTGMGGSIVK MAVESQCRAE LNKRESEWRLT 60
ALAMSAEKQA EWENKICACV AQEAPNQLTG NDVMQMLDPS TRNQALAAALT AKTVSACFKH 120
LYR 123
50 <212> Type : PRT
<211> Length : 123
SequenceName : SEQ ID 248
SequenceDescription :

Sequence
-----
55 <213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MNPLIHQAKE SSMQTRILSA VLLAFSTAAF AGGAFTLQFD NPSEDGGFTQ NQILSAPYGF 60
GCSGGNASPA LSWKNPPAGT KSFVLTVYDK DAPTGLGWMH WVADIPADV RRRNATSLQL 120
SRCASIADDQ SAAISAVISL QICRIRLTPTS YTAKMPSPCC NHANTPQSAA SAALCGTSSS 180
VSTAAA 186
60 <212> Type : PRT
<211> Length : 186
SequenceName : SEQ ID 249
SequenceDescription :

```

Sequence

<213> OrganismName : Neisseria meningitidis Z2491

<400> PreSequenceString :

```
5 MNKTLKRRVF RHTALYAAIL MFSHTGGGGG AMAQTRQYAI IMNERNQPEV QWNGSYSIKD      60
  KDRKREYTHH NHQQGGSSVS FNNSEDLVSR QSGTAVFGTA TYLPPYGVKS GFDAALAKER      120
  NNAVDWIHTT HPGLIGYSYD GVVCRSATDC PKLVYKTRFS FDNPDLAKTG GGLDKHTEPS      180
  RDNSPIYKLLK DHPWLGVSFN LGAEGIAKNG KTINKLVSSF NEKNSNNNLV YTEGRDISL      240
10 GWNQRETTAM AYYLNAKLHL LDKKQIQNIT DKTVQLGVLK PSIDVTRTNT GTAGILSYWA      300
  KWDIKDTGQI PVKLSLTQVK AGRCVNKDNF NKNTKTSSPA LTAPALWFGA QQDGKAEMYS      360
  ASVSTYPDSS SSRIFLQNLK RKTDTSRPGR YSLATLNKSD IESREPSFTS RQTVIRLDGG      420
  VQQIKLDRNN TEVTGFNGND GKNDTFGIVS EGSFMPDASE WKKVLLPWTV RAFNYDGRFN      480
  TVNKEENNGK PKYSQKYRSR NNGKHERNLG DIVNSPIVAV GEYLATSAND GMVHIFKQSG      540
  GDKRSYNLKL SYIPGTMPRK DIESKDSTLA KELRAFAEKG YVCDRYGVDG GFVLRRTTDD      600
15 QDKQKHFFMF GAMGLGGRGA YALDLTKADD NDPTKASLFD VKDNGNNGNN GNNRVELGYT      660
  VGTPIQIKTH NGKYAAFLAS GYATKQIDSG ENKTALYVYD LESNNGTLIR KIEVTDGKGG      720
  LSSPTLVDDK LDGTVDIAYA GDRGGKMYRF DLSGNNPNSW TVRTIFQGTK PITSAPAIISQ      780
  LKDKRNVVIFG TGSDDLSEDDV LSTDEQHIYG IFDNDTNTGT AQEGLGKGLL EQKLSEENKT      840
  LFLTIDYKRS DSGDGKWVVK LKDGQRVTVK PTVVLRTAFV TIHKYTGNDK CGAETAILGI      900
20 NTADGGKLTK KSARPIVPAA NSKVAQYSGD KKTSSGKSIP IGCMEKDGGT VCPNGYVYDK      960
  PVNVRYLDEK KTDGFSTTAD GDAGGSGTFK EGKKPARNNR CFSGKGVRTL LMNDLDSLDI     1020
  TGPMMCGMKRI SWREVFY                                     1037
```

<212> Type : PRT

<211> Length : 1037

```
25 SequenceName : SEQ ID 250
  SequenceDescription :
```

Sequence

<213> OrganismName : Neisseria meningitidis Z2491

<400> PreSequenceString :

```
30 MKHPKLTLLIA ALLTTAATAA PLPVVTSFSI LGDVAQKIGG ERVSIQSLVG ANQDTHAYHM      60
  TSGDIKKIRS AKLVLINGLG LEAADIQRAV KQSKVSYAEA TKGIQPLKAE EGGGHHHDHD      120
  HDHHDHDEGH BHDHGEYDPH VWNDPVLMAS YQNVAEALI KADPEGKVYY QQRLGNYQMQ      180
35 LKKLHSDAQ AFAVPAAKR KVLTHGDAFS YMCKRYHIEF IAPQVSSEA EPSAKQVAAI      240
  IRQIKREGIK AVFTENIKDT RMVDRIAKET GNVVSGKLYS DALGNAPADT YIGMYRHNIK      300
  ALTNAMKQ                                             308
```

<212> Type : PRT

<211> Length : 308

```
40 SequenceName : SEQ ID 251
  SequenceDescription :
```

Sequence

<213> OrganismName : Streptococcus pyogenes MGAS8232

<400> PreSequenceString :

```
45 MKKRILSAVL VSGVTILGAAT TVGAEDLSTK IAKQDSIISN LTTEQKAAQN QVSALQAQVS      60
  SLQSEQDKLT ARNTELEALS KRFEQETIKAL TSQIVARNEK LKNQARSAYK NNETSGYINA      120
  LLNSKSISDV VNRLVAINRA VSANAKLLEQ QKADKVSLEE KQAAQNTAIN TIAANMAMAE      180
50 ENQNTLRTQQ ANLEAATANL ALQLASATED KANLVAQKEA AEKAAAEALA QEQAQKVAQ      240
  EQAAQQAASV BAAKSAITPA PQATPAAQSS NAIEPAALTA PAAPSARFQT SYDSSNTYPV      300
  GQCTWGAKSL APWAGNNWGN GGQWYSAQA AGYRTGSTPM VGAIAVWNDG GYGHVAVVVE      360
  VQSASSIRVM ESNYSGRQYI ADHRGWFNPT GVTFIYPH                                     398
```

<212> Type : PRT

<211> Length : 398

```
55 SequenceName : SEQ ID 252
  SequenceDescription :
```

Sequence

<213> OrganismName : Streptococcus pyogenes MGAS8232

<400> PreSequenceString :

```
60 MITIKNPKIL KWLKYVLSAI LSLIILVIII GGLLFTFYIS SAPKLSEAQL KSTNSSLVYD      60
  GNNNLIADLG SEKRENTVAD SIPINLVNAI TSIEDKRFFN HRGVDLYRIF GAAFNHLSQ      120
65 TTQGGSTLDQ QLIKLAYFST NESDQTLKRK AQEVWLALQM ERKYTKQEL TFIYINKVYM      180
  NGNYGMLTAA KSYYGKDLKD LSYAQLALLA GIPQAPSQYD PYLHPEAAQN RRNVVLQQMY      240
  MEKHLTKAEY ETAIATPVAE GLQSLQQRST YPKYMDNYLK QVIEEVKKET NKDIFTAGLK      300
```

VYTNIIPDAQ QTLYNIYHSG DYVYYPDQDF QVASTIVDVT NGHVIAQLGG RNQDENVSFG 360
TNQAVLTDRD WGSTMKPITA YAPAIESGVY TSTAQSTNDS VYYWPGTTTQ LFNWDLRYNG 420
WMTIQAAIML SRNVPVAVRAL EAAGLDYARS FLSSLGINYP EMHYSNAISS NNSSSDKKYG 480
ASSEKMAAAY AAFANGGIYH KPRYVKNVEF SDGTSKTFDE KGKRAMKETT AYMTDMLKT 540
5 VLTYGTGTAA AIPGVAQAGK TGTSNYTDEE LAKIGEKYGL YPDYVGTLAP DENFVGFTKR 600
YAMAVWTGYK NRLTPVYGSS LEIASDVYRS MMTYLTNGYS EDWTMPNGLY RSGGFLYLSG 660
TYASNTDYTN SVYNNLYSNM TTTASSQTTS DDTSSSNDTS NSTNTDNNGS HPSTDDKKT 720
H 721
<212> Type : PRT
10 <211> Length : 721
SequenceName : SEQ ID 253
SequenceDescription :

Sequence
15 -----
<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
MIITKKSILFV TSVALSLAPL VTAQAQEWTP RSVTEIKSEL VLVDNVFTYT VKYGDTLSTI 60
AEAMGIDVHV LGDINHIANI DLIFPDILT ANYNQHQGAT TLTVQAPASS PASVSHVPSS 120
20 EPLPQASATS QSTVPMAPSA TPSDVPTTPL ASAKPDSFVT ASSELTSSSTN DVSTELSSSES 180
QKQPEVSQEA VPTPKAAETT EVEPKTDISE DPTSANRPVP NESASEEASS AAPAQAPAEK 240
EETSQMLTAP AAQKAVADTT SVATSNGLSY APNHAYNPMN AGLQPQTAAF KEEVASAFGI 300
TSFSGYRPGD PGDHGKGLAI DFMVPVSSTL GDQVAQYAIH HMAERGISYV IWKQRFYAPF 360
ASIYGPAYTW NPMFDRGSIT ENHYDHVHVS FNA 393
25 <212> Type : PRT
<211> Length : 393
SequenceName : SEQ ID 254
SequenceDescription :

Sequence
30 -----
<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
MKKKILLMMS LISVFFAWQL TQAKQVLAEG KVKVVTTFYP VYEFTKGIVG NDGDVSMMLK 60
35 AGTEPHDFEP STKDIKKIQD ADAFVYMDN METWVSDVKK SLTSKKVTIV KGTGNMLLVA 120
GAGHDHHHED ADKKHEHNKH SEEGHNHAFD PHVWLSPYRS ITVVENIRDS LSKAYPEKAE 180
NFKANAATYI EKLKELDKDY TAALSDAKQK SFVTQHAAFG YMALDYGLNQ ISINGVTPDA 240
EPSAKRIATL SKYVKKYGIK YIYFEENASS KVAKTLAKEA GVKA AVLSP L EGLTKKEMKA 300
GQDYFTVMRK NLETLRLTTD VAGKEILPEK DTTKTVYNGY FKDKVVKDRQ LSDWSGWSQS 360
40 VYPYLQDGT L DQVWDYKAKK SKGKMTAAEY KDYYTTGYKT DVEQIKINGK KKTMTFVRNG 420
BKKTFTYTYA GKEILTYPKG NRGVRFMF EA KEPNAGEFKY VQFSDHAIAP EKAHFHLYW 480
GGDSQEK LHK ELEHWPTYG SDLSGREIAQ EINAH 515
<212> Type : PRT
<211> Length : 515
45 SequenceName : SEQ ID 255
SequenceDescription :

Sequence
50 -----
<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
MKKFHRFLVS GVILLGFNG L VPTMPSTLIS QQENLVHAAV LGDNYPSKWK KGNIDSWNM 60
YIRQCTSF AA FRLSSANGFQ L PKGYGNACT WGHI AKNQGY PVNKTPSIGA IAWFDKNAYQ 120
SNAAYDHVAV VADIRGDTV T IEEYNYNAGQ GPERYHKRQI PKSQVSGYIH FKDLSSQTSH 180
55 SYPRQLKHIS QASFDPSGT Y HFTTRLPVKG QTSIDSPDLA YYEAGQSVYV DKVVV TAGGYT 240
WLSYLSFSGN RRYIPIKEPA QSVVQNDNTK PSIKVGDTV T FPGVFRVDQL VNNLIVNKEL 300
AGGDPTPLNW IDPTPLDET D NQGVKVLGNQI LRVGEYFTVT GSYKVLKIDQ PSNGIYVQIG 360
SRGTWVNADK ANKL 374
<212> Type : PRT
60 <211> Length : 374
SequenceName : SEQ ID 256
SequenceDescription :

Sequence
65 -----
<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :

MLKFTSNILA TSVAEETTQVA PGGCCCCCTT CCFSIATGSG NSQGGSGSYT PGK

<212> Type : PRT

<211> Length : 53

5 SequenceName : SEQ ID 257
 SequenceDescription :

Sequence

10 <213> OrganismName : Streptococcus pyogenes MGAS8232

<400> PreSequenceString :

MGESYSVEAV	LTAVDKTFGK	TLQSAIRSIE	GLEKRSTGFS	SVSQKASSMF	KSMLGANLAG	60
QAISAMTRTV	SSGLGSMLGE	MNSSAKAWKT	FDANLADIGF	GKKQILAVKT	AMQDYATKTI	120
YSASDMASTY	AQLAAVGVD	TGKLVKAFGG	LAASAENPKQ	AMKSISQQMT	QAVGRPTVAV	180
15 QDFRIMLEQT	PAGMAKVAKS	MGKNLDELVA	DIQAGRVKTS	DFLEAVKKAG	NDKSFQKMAT	240
EFKTVDAQID	GMREGLSNKL	QPAFEKVNQF	GIRAIEAIGK	QLDKVDFSKF	ASNLGKFLEG	300
INIDKIVSNI	SSAVSSVTISK	VKEFWDGFKQ	TGAISAFSGA	LQSVWGALKN	VASAMSGGNW	360
KTFGATVGGI	VKHVSNFAKA	VSDVLGKMDP	GRLRSWIATF	AAVAGGFKLF	EKLTGQSVIG	420
SPLDKIGSKF	GLFGNKAKEG	TDKASNGARR	SGGIISQIFS	GLGNIVKSAG	TAISTAAGKI	480
20 GVGIKTALSG	IPPYH					495

<212> Type : PRT

<211> Length : 495

SequenceName : SEQ ID 258

SequenceDescription :

Sequence

25 <213> OrganismName : Streptococcus pyogenes MGAS8232

<400> PreSequenceString :

30 MKKGFFLMVM	VVSLVMIAGC	DKSANPKQPT	QGMSVVSIFY	PMYAMTKEVS	GDLNDVRMIQ	60
SGAGIHSFEP	SVNDVAAIYD	ADLFVYHSHT	LEAWARLDLP	NLKKSKVDVF	EASKPLTLDR	120
VKGLEDEMEVT	QGIDPATLYD	PHTWTDPVLA	GEEAVNIAKE	LGRLDPKHKD	SYTKNAKAFK	180
KEAEQLTEBY	TQKFKKVRSK	TFVTQHTAFS	YLAKRFGLKQ	LGISGISPEQ	EPSPRQLKEI	240
QDFVKEYNVK	TIFAEQNVNP	KIAHAIAKST	GAKVKTLSPL	EAAPSGNKTY	LENLRANLEV	300
35 LYQQLK						306

<212> Type : PRT

<211> Length : 306

SequenceName : SEQ ID 259

SequenceDescription :

Sequence

40 <213> OrganismName : Streptococcus pyogenes MGAS8232

<400> PreSequenceString :

45 MEKKQRFSLR	KYKSGTFSVL	IGSVFLMMTT	TVAADELSTM	SEPTITNHTQ	QQAQHLTNTE	60
LSSAESKSD	TSQITPKTNR	EKEQPQGLVS	EPTTTELADT	DAAPMANTGP	DATQKSASLP	120
PVNTDVIDHW	KTGAWDKGY	KGQGVVAVI	DTGIDPAHQS	MRISDVSTAK	VKSKEDMLAR	180
QKAAGINYGS	WINDKVFFAH	NYVENSNDIK	ENQFEDFDED	WENFEFDAEA	EPKAIKHKHI	240
YRPQSTQAPK	ETVIKTEETD	GSHDIDWTQT	DDDTKYESHG	MHVTGIVAGN	SKEAAATGER	300
50 FLGIAPEAQV	MFMRVVFANDV	MGSAESLFIK	AIEDAVALGA	DVINLSLGT	NGAQLSGSKP	360
LMEAIEKAKK	AGVSVVVAAG	NERVYGSDDH	DPLAINPDYG	LVGSPSTGRT	PTSVAAINSK	420
WVIQRLMTVK	ELENRADLNH	GKAIYSESVD	FKNIKDSLGY	DKSHQFAYVK	ESTDAGYKAQ	480
DVKDKIALIE	RDPNKTYDEM	IALAKKHGAL	GVLIFFNNKPG	QSNRSMRLTA	NGMGIPSAFI	540
SHEFGKAMSQ	LNGNGTGSLE	FDSVVSAPKS	QKGNEMNHFS	NWGLTSDGYL	KPDITAPGGD	600
55 IYSTYNDNHY	GSQGTGSMAS	PQIAGASLLV	KQYLEKTQPN	LPKEKIADIV	KNLLMSNAQI	660
HVNPETKTTT	SPRQQAGALL	NIDGAVTSGL	YVTGKDNYS	ISLGNITDTM	TFDVTVHNLS	720
NKDKTLRYDT	ELLTDHVDPO	KGRFTLTSRS	LKTYQGGEVT	VPANGKVTVR	VTMDVSQFTK	780
ELTKQMSNGY	YLEGFVRFRD	SQDDQLNRVN	IPFVFGKGQF	ENLAVAEESE	YRLKSQKGTG	840
FYFDESGPKD	DIYVGKHFTG	LVTLGSETNV	STKTISDNGL	HTLGTGKNAD	GKFILEKNAQ	900
60 GNPVLATSPN	GDNNQDFAAF	KGVFLRKYQG	LKASVYHASD	KEHKNPLWVS	PESFKGDKNF	960
NSDIRFAKST	TLLGTAFSGK	SLTGAELPDG	YYHYVVSYP	DVVGAKRQEM	TFDMILDROK	1020
PVLSQATFDP	ETNRFKPEPL	KDRGLAGVRK	DSVFYLERKD	NKPYTVTIND	SYKYVSVEDN	1080
KTFVERQADG	SFILPLDKAK	LGDFYYMVED	FAGNVAIAKL	GDHLPQTLGK	TPIKLKLTDG	1140
NYQTKETLKD	NLEMTQSDTG	LVTNQAQLAV	VHRNQPSQL	TKMNQDFFIS	PNEDGNKDFV	1200
65 AFKGLKNVY	NDLTVNVYAK	DDHQKQTPIW	SSQAGASASA	IESTAWYGIT	ARGSKVMPGD	1260
YQYVVVYRDE	HGKEHQKQYT	ISVNDKKPMI	TQGRFDTING	VDHFTPDKTK	ALGSSGIVRE	1320
EVFYLAKKNG	RKFDVTEGKD	GITVSDNKMY	IPKNPDGSYT	ISKRDGVTLS	DYYYYLVEDRA	1380

GNVSFATLRD LKAVGKDKAV VNFGLDLPVP EDKQIVNFTY LVRDADGKPI ENLEYNNNSG 1440
NSLILPYGKY TVELLTYDTN AAKLESDDKIV SFTLSADNNF QQVTFKMTML ATSQITAHFD 1500
HLLPEGSRVLS LKTAQGQLIP LEQSLYVPKA YGKTVQEGTY EVVVSPLPKGY RIEGNTKVNT 1560
LPNVEHELSSL RLVKVGDAASD STGDHKVMSK NNSQALTAFA TPTKTTTSAI AKALPSAGEK 1620
5 MGLKLRIVGL VLLGLTLCVFS RKKSTKD 1647
<212> Type : PRT
<211> Length : 1647
SequenceName : SEQ ID 260
SequenceDescription :

10 Sequence

<213> OrganismName : Treponema pallidum subsp. pallidum str. Nichols
<400> PreSequenceString :
15 MMRSLSFGVS GMQNHQTRMD VIGNNVANVN TTGFKRGRVN FQDLISQQLS AAARPNEEVG 60
GVNPKFVGLG VLIASIDTVH TQCALQTTGI NTDVSIQSGG FFVLKSSEKI FFTRAGAFGV 120
DNAGTLVNPA NGMRVQGWMA QDVAGERLIN SSAQTQDLVI PIGQKIDAQQ TSTVHYACNL 180
DKRLPELAAD ANEADVRKST WTTDFQVYDS FGQHTLQIN FSRVPGTNNQ WQATVAVDPG 240
TEVDTQTRVG VGTSDGAANT FIVNFDNFGH LASVTDTAGN VTGPTQVLL EASYDVVGAN 300
20 PDDAGQVTRH AFTLNLGEIG TARNTTQFA ERSTTKAYRQ DGYAMGYLEN FKIDQSGVIT 360
GVSYNGVSQD IGQLALAGFA NQGGLEKAGE NTYVQSNNSG IANISTSGVM GKGLIAGTL 420
EMSNVDLTDQ FTDMIITQKG FOAGAKTIQT SDTMLDTVLS LKR 463
<212> Type : PRT
<211> Length : 463
25 SequenceName : SEQ ID 261
SequenceDescription :

Sequence

30 <213> OrganismName : Treponema pallidum subsp. pallidum str. Nichols
<400> PreSequenceString :
MGCMRWGSVL CVVVGVGASG GVLGQEFSPK LTGSATLEWG ISYGKGVGSH QAPGAVMGT 60
GPYNLKHGFR TTNVGVGFPP LVMRTTHTRR QHPALYAEI KVADLQADLS QGKAGFAVKR 120
KGKVEATLHC YGAYLTIGKN PTFLTNFARL WKPWVTAQYQ EDAVQYAPGF GGLGGKVGYS 180
35 AQDIGGSGVS LDVGFLSFAS NGAWDSTDPT HSKYGFADL KLMYARAGHP LCTVELASNV 240
TLEDGYLIGA QKDANNQNKD KLLWNVGGRL TLEPGAGFRF SFALDAGNQH QSAQDFQNR 300
QRAQSELTAL SNNLFQGESQ KQEAWVTQVV QATQTVTAG VRSALSRGT TYINALEAVQ 360
PNPAKPTGKV VQNLHTPQGS PPNLPLPAL PAFSLMGQVL LQYDAEQVVK GFEQVQTQIV 420
TEINQKVQAA VAKNNANMQA VGGSLGDTAR MVGEALIKQQ LSRKQNSILT MVSQDEVKQ 480
40 DLADLVPMR TEITAFFASV QQHITTEVKK KTDALNAGQQ IRQATQNLRA SAWRAFLMGV 540
SAVCLYLDY NVAFDALFTA QWKWLSGGYI FATAPANVFG TRVLDNTIAS CGDFAGFLKL 600
ETKSGDPYTY LLTGLDAGVE TRVYIPLTHD LYKNNNGNPL PSGGSSGHIG LPVVGKAWCS 660
YRIPVQDYGW VKPSVTVHAS TNRAHLNAPA AGGAVGATYL TKEYCAQLRA GISASLIEKT 720
VFSLDWEQGM LSDVPYLLVS ECLTQIGIRI VCGVTLWS 758
45 <212> Type : PRT
<211> Length : 758
SequenceName : SEQ ID 262
SequenceDescription :

50 Sequence

<213> OrganismName : Treponema pallidum subsp. pallidum str. Nichols
<400> PreSequenceString :
MGRQVMQAGV LAGMVCAASG YAGVLTQVVS GTAQLQWGIA FQKNPRTGPG KHTHGFRITN 60
55 SLTISLPLVS KHTHTRRGEA RSGVWAQLQL KDALVELASS KSSTALSFTK PTASFQATLH 120
CYGAYLTVGT SPSCVNVFAQ LWKPFVTRAY SEKOTRYAPG FSGSGAKLGY QAHNVGNSGV 180
DVDIGFLSFL SNGAWDSTD TSKYGFADL ATLSYGVDRQ RLLTLELAGN ATLDQNYVKG 240
TEDSKNENKT ALLWGVGGRL TLEPGAGFRF SFALDAGNQH QSNAHAQTQIE RAILKAREVF 300
RRVEGKLVQN LPNIMPPGI TEQTTLIEMV GLAALIAEGT LGSATQTVLA AGALAALVSQ 360
60 LVPNIEQGVV DVFRSSDPRV VTAKLLAFLE RAPMNALNID ALLRMQWKWL SSGIYFATAG 420
TNIFGKRVFA PTRAHYFDFA GFLKLETKSG DPYTHLLTGL NAGVEARVYI PLTYIRYRN 480
GGYELNGAVP TGTINMPILG KAWCSYRIPL GSHAWLAPHT SVLGTNRFN IINPAGNLLN 540
ERALQYQVGL TFSPEKVEL SAQWEQGVLA DAPYMGIAES IWSERHFGTL VCGMKVTW 598
65 <212> Type : PRT
<211> Length : 598
SequenceName : SEQ ID 263

SequenceDescription :

Sequence

5 <213> OrganismName : Treponema pallidum subsp. pallidum str. Nichols
 <400> PreSequenceString :
 MGRQVMQAGV LAGMVCAASG YAGVLTPOVS GTAQLQWGIA FQKNPRTGPG KHTHGFRRTN 60
 SLTISLPLVS KHTHTRRGEA RSGVWAQLQL KDLAVELASS KSSTALSFTK PTASFQATLH 120
 CYGAYLTVGT SPSCVNVFAQ LWKPFVTRAY SEKDRYAPG FSGSGAKLGY QAHNVGNSGV 180
 10 DVIDGFLSFL SNGAWDSTDT THSKYGFAD ATLSYGVDRQ RLLTLELAGN ATLDQNYVKG 240
 TEDSKNENKT ALLWGVGGRL TLEPGAGFRF SFALDAGNQH QSNAHAQTQE RAILKAREVF 300
 RRVGKLVQN LPNIMPPGI TEQTTLIEMV GLAALIAEGT LGSATQTVLA AGALAALVSQ 360
 LVPNIEQGVV DFRSSDPRV VTAKLLAFLE RAPMNALNID ALLRMQWKWL SSGIYFATAG 420
 TNIFGKRVFA TTRAHYFDEA GFLKLETKSG DPYTHLLTGL NAGVEARVYI PLTYIRYRNN 480
 15 GGYELNGAVP PGTINMPLG KAWCSYRIPL GSHAWLAPHT SVLGTNRNFN IINPAGNLLN 540
 ERALQYQVGL TSPFEKVEL SAQWEQGVLA DAPYMGIAES IWSEHFGTL VCGMKVTW 598

<212> Type : PRT

<211> Length : 598

20 SequenceName : SEQ ID 264
 SequenceDescription :

Sequence

25 <213> OrganismName : SARS coronavirus Frankfurt 1
 <400> PreSequenceString :
 MFIFLLFLTL TSGSDDLDRCT TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL 60
 PFYSNVTGFH TINHTFGNPV IPFKDGIYFA ATEKSNVVRG WVFGSTMNNK SQSVIIINNS 120
 TNVIRACNF ELCDNPFFAV SKPMGTQHT MIFDANFNCT FEYISDAFSL DVSEKSGNFK 180
 30 HLREFVFKNK DGFLYVYKGY QPIDVVRDLP SGFNTLKPIF KLPLGINITN FRAILTAFSP 240
 AQDIWGTSA AYPVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY 300
 QTSNFRVVPV GDVVRFPNIT NLCPFGEVFN ATKFPVSVAW ERKKISNCVA DYSLVYNSTF 360
 FSTFKCYGVS ATKLNLCFS NVYADSFVVK GDDVRQIAPG QTGVADYNY KLPDDFMGCV 420
 LAWNTRNIDA TSTGNYNYKY RYLRHGKLRP FERDISNVPF SPDGKPCPP ALNCYWPLND 480
 35 YGYTTTGIG YQPYRVVLS FELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLTP 540
 SSKRFQPFQ FGRDVSDFTD SVRDPKTSEI LDISPCSFSG VSVITPGTNA SSEVAVLYQD 600
 VNCTDVSTAI HADQLTPAWR IYSTGNVVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY 660
 HTVSLRSTS QKSIVAYTMS LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC 720
 NMYICGDSTE CANLLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKYFG 780
 40 GFNFQILPD PLKPTKRSFI EDLLFNKVTI ADAGFMKQYG ECLGDINARD LICAQKFNGL 840
 TVLPPLLTDD MIAAYTAALV SGTATAGWTF GAGAAQIPF AMQMAYRFG IGVTQNVLYE 900
 NQKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAIISSVLN 960
 DILSRDLKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK 1020
 RVDFCGKGYH LMSFPQAAPH GVVFLHVTVV PSQERNFTTA PAICHEGKAY FPREGVVFVN 1080
 45 GTSWFITQRN FFSPQIITD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN 1140
 HTSPDVDFGD ISGINASVFN IQKEIDRLNE VAKNLNESLI DLQELGKYEQ YIKWPWYVWL 1200
 GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCCKFDE DDSEPVKGV KLHYT 1255

<212> Type : PRT

<211> Length : 1255

SequenceName : SEQ ID 265

SequenceDescription :

Sequence

55 <213> OrganismName : SARS coronavirus HSR 1
 <400> PreSequenceString :
 MFIFLLFLTL TSGSDDLDRCT TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL 60
 PFYSNVTGFH TINHTFGNPV IPFKDGIYFA ATEKSNVVRG WVFGSTMNNK SQSVIIINNS 120
 60 TNVIRACNF ELCDNPFFAV SKPMGTQHT MIFDANFNCT FEYISDAFSL DVSEKSGNFK 180
 HLREFVFKNK DGFLYVYKGY QPIDVVRDLP SGFNTLKPIF KLPLGINITN FRAILTAFSP 240
 AQDIWGTSA AYPVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY 300
 QTSNFRVVPV GDVVRFPNIT NLCPFGEVFN ATKFPVSVAW ERKKISNCVA DYSLVYNSTF 360
 FSTFKCYGVS ATKLNLCFS NVYADSFVVK GDDVRQIAPG QTGVADYNY KLPDDFMGCV 420
 65 LAWNTRNIDA TSTGNYNYKY RYLRHGKLRP FERDISNVPF SPDGKPCPP ALNCYWPLND 480
 YGYTTTGIG YQPYRVVLS FELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLTP 540
 SSKRFQPFQ FGRDVSDFTD SVRDPKTSEI LDISPCSFSG VSVITPGTNA SSEVAVLYQD 600

5 VNCTDVSTAI HADQLTPAWR IYSTGNNVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY 660
HTVSLLRSTS QKSIVAYTMS LGADSSIAYS NNTTIAIPTNF SISITTEVMP VSMAKTSVDC 720
NMYICGDSTE CANLLQLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKYFG 780
GFNFSQLLPD PLKPTKRSFI EDLLFNKVTI ADAGFMKQYG ECLGDINARD LICAQKFNGI 840
TVLPPLLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTQNVLYE 900
10 NQKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN 960
DILSRDLKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK 1020
RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVVFVN 1080
GTSWFITQRN FFSPQIITTD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN 1140
10 HTSPDVLGD ISGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYEY YIKWPWYVWL 1200
GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCCKFDE DDSEPVLKGV KLHYT 1255

<212> Type : PRT
<211> Length : 1255
15 SequenceName : SEQ ID 266
SequenceDescription :

Sequence

20 <213> OrganismName : SARS coronavirus ZJ01
<400> PreSequenceString :
MFIFLLFLTL TSGSDLDRCI TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL 60
PFYSNVTGFH TINHTFGNPV IPFKDGIYFA ATEKSNVVRG WVFGSTMNNK SQSVIIINNS 120
TNVVIRACNF ELCNDPFFAV SKPMGTQTHI MIFDIAFNCT FEYISDAFSL DVSEKSGNFK 180
25 HLREFVFKNK DGFLYVYKGY QPIDVVRDLI SGFNTLKPIF KLPLGINITN FRAILTAFSP 240
AQDIWGTSA AAYFVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY 300
QTSNFRVVPV GDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKKISNCVA DYSLVYNSTF 360
FSTFKCYGVS ATKLNLCFS NVYADSFVVK GDDVRQIAPG QTGVIADYNY KLPDDFMGCV 420
LAWNTRNIDA TSTGNVNYKY RYLRHGKLRP FERDISNVPI SPDGKPCPTP ALNCYWPLND 480
30 YGFYTTTGIG YQPYRVVLS FELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLT 540
SSKRFQPFQ FGRDVSDFD SVRDPKTSEI LDISPCSFGG VSVITPGTNA SSEVAVLYQD 600
VNCTDVSTAI HADQLTPAWR IYSTGNNVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY 660
HTVSLLRSTS QKSIVAYTMS LGADSSIAYS NNTTIAIPTNF SISITTEVMP VSMAKTSVDC 720
NMYICGDSTE CANLLQLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKYFG 780
35 GFNFSQLLPD PLKPTKRSFI EDLLFNKVTI ADAGFMKQYG ECLGDINARD LICAQKFNGI 840
TVLPPLLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTQNVLYE 900
NQKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN 960
DILSRDLKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK 1020
RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVVFVN 1080
40 GTSWFITQRN FFSPQIITTD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN 1140
HTSPDVLGD ISGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYEY YIKWPWYVWL 1200
GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCCKFDE DDSEPVLKGV KLHYT 1255

<212> Type : PRT
45 <211> Length : 1255
SequenceName : SEQ ID 267
SequenceDescription :

Sequence

50 <213> OrganismName : SARS coronavirus TW1
<400> PreSequenceString :
MFIFLLFLTL TSGSDLDRCI TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL 60
PFYSNVTGFH TINHTFGNPV IPFKDGIYFA ATEKSNVVRG WVFGSTMNNK SQSVIIINNS 120
55 TNVVIRACNF ELCNDPFFAV SKPMGTQTHI MIFDIAFNCT FEYISDAFSL DVSEKSGNFK 180
HLREFVFKNK DGFLYVYKGY QPIDVVRDLI SGFNTLKPIF KLPLGINITN FRAILTAFSP 240
AQDIWGTSA AAYFVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY 300
QTSNFRVVPV GDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKKISNCVA DYSLVYNSTF 360
FSTFKCYGVS ATKLNLCFS NVYADSFVVK GDDVRQIAPG QTGVIADYNY KLPDDFMGCV 420
60 LAWNTRNIDA TSTGNVNYKY RYLRHGKLRP FERDISNVPI SPDGKPCPTP ALNCYWPLND 480
YGFYTTTGIG YQPYRVVLS FELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLT 540
SSKRFQPFQ FGRDVSDFD SVRDPKTSEI LDISPCSFGG VSVITPGTNA SSEVAVLYQD 600
VNCTDVSTAI HADQLTPAWR IYSTGNNVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY 660
HTVSLLRSTS QKSIVAYTMS LGADSSIAYS NNTTIAIPTNF SISITTEVMP VSMAKTSVDC 720
65 NMYICGDSTE CANLLQLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKYFG 780
GFNFSQLLPD PLKPTKRSFI EDLLFNKVTI ADAGFMKQYG ECLGDINARD LICAQKFNGI 840
TVLPPLLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTQNVLYE 900

NQKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN 960
 DILSRLDKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK 1020
 RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVVFVN 1080
 GTSWFTTQRN FSPQIITTD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN 1140
 5 HTSPDVLGD ISGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYEQ YIKWPWYVWL 1200
 GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCKKFDE DDSEPVKGV KLHYT 1255

<212> Type : PRT
 <211> Length : 1255
 10 SequenceName : SEQ ID 268
 SequenceDescription :

Sequence

15 <213> OrganismName : SARS coronavirus CUHK-Su10
 <400> PreSequenceString :

MFIFLLFLTL TSGSDLDRCT TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL 60
 PFYSNVTGFH TINHTFGNPV IPFKDGIYFA ATEKSNVVRG WVFGSTMNNK SQSVIIINNS 120
 TNVVIRACNF ELCNDPFFAV SKPMGTQHT MIFDNAFNCT FEYISDAFSL DVSEKSGNFK 180
 20 HLREFVFNK DGLYVYKGY QPIDVVRDLP SGFNTLKPIF KLPLGINITN FRAILTAFSP 240
 AQDIWGTSA AYPVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY 300
 QTSNFRVVPV GDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKKISNCVA DYSVLNSTF 360
 FSTFKCYGVS ATKLNLCFS NVYADSFVVK GDDVRQIAPG QTGVIADYNY KLPDDFMGCV 420
 LAWNTRNIDA TSTGNYNKY RYLRHGKLRP FERDISNVPF SPDGKPCPTP ALNCYWPLND 480
 25 YGFYTTTGIG YQPYRVVLS FELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLTP 540
 SSKRFQPFQ FGRDVSDFTD SVRDPKTSEI LDISPCSFEG VSVITPGTNA SSEVAVLYQD 600
 VNCTDVSTAI HADQLTPAWR IYSTGNVVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY 660
 HTVSLRSTS QKSIVAYTMS LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC 720
 NMYICGDSTE CANLLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKYFG 780
 30 GFNFSQLPD PLKPTKRSFI EDLLFNKVTI ADAGFMKQYG ECLGDINARD LICAQKFNGL 840
 TVLPPLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTQNVLYE 900
 NQKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN 960
 DILSRLDKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK 1020
 RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVVFVN 1080
 35 GTSWFTTQRN FSPQIITTD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN 1140
 HTSPDVLGD ISGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYEQ YIKWPWYVWL 1200
 GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCKKFDE DDSEPVKGV KLHYT 1255

<212> Type : PRT
 40 <211> Length : 1255
 SequenceName : SEQ ID 269
 SequenceDescription :

Sequence

45 <213> OrganismName : SARS coronavirus Urbani
 <400> PreSequenceString :

MFIFLLFLTL TSGSDLDRCT TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL 60
 PFYSNVTGFH TINHTFGNPV IPFKDGIYFA ATEKSNVVRG WVFGSTMNNK SQSVIIINNS 120
 50 TNVVIRACNF ELCNDPFFAV SKPMGTQHT MIFDNAFNCT FEYISDAFSL DVSEKSGNFK 180
 HLREFVFNK DGLYVYKGY QPIDVVRDLP SGFNTLKPIF KLPLGINITN FRAILTAFSP 240
 AQDIWGTSA AYPVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY 300
 QTSNFRVVPV GDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKKISNCVA DYSVLNSTF 360
 FSTFKCYGVS ATKLNLCFS NVYADSFVVK GDDVRQIAPG QTGVIADYNY KLPDDFMGCV 420
 55 LAWNTRNIDA TSTGNYNKY RYLRHGKLRP FERDISNVPF SPDGKPCPTP ALNCYWPLND 480
 YGFYTTTGIG YQPYRVVLS FELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLTP 540
 SSKRFQPFQ FGRDVSDFTD SVRDPKTSEI LDISPCSFEG VSVITPGTNA SSEVAVLYQD 600
 VNCTDVSTAI HADQLTPAWR IYSTGNVVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY 660
 HTVSLRSTS QKSIVAYTMS LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC 720
 60 NMYICGDSTE CANLLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKYFG 780
 GFNFSQLPD PLKPTKRSFI EDLLFNKVTI ADAGFMKQYG ECLGDINARD LICAQKFNGL 840
 TVLPPLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTQNVLYE 900
 NQKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN 960
 DILSRLDKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK 1020
 RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVVFVN 1080
 65 GTSWFTTQRN FSPQIITTD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN 1140
 HTSPDVLGD ISGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYEQ YIKWPWYVWL 1200

GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCCKFDE DDSEPVLKGV KLHYT 1255

<212> Type : PRT
<211> Length : 1255
5 SequenceName : SEQ ID 270
 SequenceDescription :

Sequence

10 <213> OrganismName : SARS coronavirus
 <400> PreSequenceString :
MFIFLLFLTL TSGSDLDRC TFDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL 60
PFYSNVTFGH TINHTFGNPV IPFKDGIYFA ATEKSNVVRG WVFGSTMNNK SQSVIIINNS 120
TNVVIRACNF ELCNPPFFAV SKPMGTQHT MIFDNAFNCT FEYISDAFSL DVSEKSGNFK 180
15 HLREFVFNK DGFLYVYKGY QPIDVVRDL SGFNTLKPIF KLPLGINITN FRAILTAFSP 240
AQDIWGTSA AYFVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY 300
QTSNFRVVP GDVVRFPNIT NLCPFGEVFN ATKFPVSYAW ERKKISNCVA DYSVLYNSTF 360
FSTFKCYGVS ATKLNDLCFS NVYADSFVVK GDDVRQIAPG QTGVADYNY KLPDDFMGCV 420
LAWNTRNIDA TSTGNYNYKY RYLRHGKLRP FERDISNVPF SPDGKPCPTP ALNCYWPLND 480
20 YGFYTTTGIG YQPYRVVVL FELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLT 540
SSKRFQPFQ FGRDVSDFTD SVRDPKTSEI LDISPCAFGG VSVITPGTNA SSEVAVLYQD 600
VNCTDVSTAI HADQLTPAWR IYSTGNNVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY 660
HTVSLRSTS QKSIVAYTMS LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC 720
NMYICGDSTE CANLLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKYFG 780
25 GFNFSQLPD PLKPTKRSFI EDLLFNKVTI ADAGFMKQYG ECLGDINARD LICAQKFNG 840
TVLPPLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTQNVLYE 900
NOKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN 960
DILSRDLKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK 1020
RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVVFVN 1080
30 GTSWFITQRN FFSQIITTD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN 1140
HTSPDVLGD ISGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYEY YIKWPWYVWL 1200
GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCCKFDE DDSEPVLKGV KLHYT 1255

<212> Type : PRT
35 <211> Length : 1255
 SequenceName : SEQ ID 271
 SequenceDescription :

Sequence

40 <213> OrganismName : SARS coronavirus Tor2
 <400> PreSequenceString :
MFIFLLFLTL TSGSDLDRC TFDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL 60
PFYSNVTFGH TINHTFGNPV IPFKDGIYFA ATEKSNVVRG WVFGSTMNNK SQSVIIINNS 120
45 TNVVIRACNF ELCNPPFFAV SKPMGTQHT MIFDNAFNCT FEYISDAFSL DVSEKSGNFK 180
HLREFVFNK DGFLYVYKGY QPIDVVRDL SGFNTLKPIF KLPLGINITN FRAILTAFSP 240
AQDIWGTSA AYFVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY 300
QTSNFRVVP GDVVRFPNIT NLCPFGEVFN ATKFPVSYAW ERKKISNCVA DYSVLYNSTF 360
FSTFKCYGVS ATKLNDLCFS NVYADSFVVK GDDVRQIAPG QTGVADYNY KLPDDFMGCV 420
50 LAWNTRNIDA TSTGNYNYKY RYLRHGKLRP FERDISNVPF SPDGKPCPTP ALNCYWPLND 480
YGFYTTTGIG YQPYRVVVL FELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLT 540
SSKRFQPFQ FGRDVSDFTD SVRDPKTSEI LDISPCAFGG VSVITPGTNA SSEVAVLYQD 600
VNCTDVSTAI HADQLTPAWR IYSTGNNVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY 660
HTVSLRSTS QKSIVAYTMS LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC 720
55 NMYICGDSTE CANLLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKYFG 780
GFNFSQLPD PLKPTKRSFI EDLLFNKVTI ADAGFMKQYG ECLGDINARD LICAQKFNG 840
TVLPPLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTQNVLYE 900
NOKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN 960
DILSRDLKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK 1020
60 RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVVFVN 1080
GTSWFITQRN FFSQIITTD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN 1140
HTSPDVLGD ISGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYEY YIKWPWYVWL 1200
GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCCKFDE DDSEPVLKGV KLHYT 1255

65 <212> Type : PRT
 <211> Length : 1255
 SequenceName : SEQ ID 272

SequenceDescription :

Sequence

5 <213> OrganismName : SARS coronavirus GD01
<400> PreSequenceString :
MFIFLLFLTL TSGSDLDRCT TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL 60
PFYSNVITGFH TINHTFDNPV IPFKDGIYFA ATEKSNVVRG WVFGSTMNNK SQSVIIINNS 120
TNVIVIRACNF ELCNDPFFAV SKPMGTQTHT MIFDNAFNCT FEYISDAFSL DVSEKSGNFK 180
10 HLREFVFKNK DGFLYVYKGY QPIDVVRDLP SGFNTLKPIF KLPLGINITN FRAILTAFLP 240
AQDTWGTSA AAYFVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY 300
QTSNFRVVPV RDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKRISNCVA DYSLVYNSTF 360
FSTFKCYGVS ATKLNLCFVS NVYADSFVVK GDDVRQIAPG QTGVIADYNY KLPDDFMGCV 420
LAWNTRNIDA TSTGNYNYKY RYLRHGKLRP FERDISNVPF SPDGKPCPTP ALNCYWPLND 480
15 YGFYTTTGIG YQPYRVVVLV YELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLTP 540
SSKRFQPFQ FGRDVSDFTD SVRDPKTSEI LDISPCSFSG VSVITPGTNA SSEVAVLYQD 600
VNCTDVSTAI HADQLTPAW IYSTGNNVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY 660
HTVSLLRSTS QKSIVAYTMS LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC 720
NMYICGDSTE CANLLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKDFG 780
20 GFNFSQLPDP PLKSTKRSFI EDLLFNKVTI ADAGFMKQYG ECLGDINARD LICAQKFNGI 840
TVLPPLLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTONVLYE 900
NQKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN 960
DILSRDLKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK 1020
RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVVFVN 1080
25 GTSWFITQRN FFSPPQIITD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN 1140
HTSPDVLGD ISGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYEY YIKWPWYVWL 1200
GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCKCFDE DDSEPVKGV KLHYT 1255

<212> Type : PRT

30 <211> Length : 1255

SequenceName : SEQ ID 273

SequenceDescription :

Sequence

35 <213> OrganismName : SARS coronavirus CUHK-W1
<400> PreSequenceString :
MFIFLLFLTL TSGSDLDRCT TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL 60
PFYSNVITGFH TINHTFDNPV IPFKDGIYFA ATEKSNVVRG WVFGSTMNNK SQSVIIINNS 120
40 TNVIVIRACNF ELCNDPFFAV SKPMGTQTHT MIFDNAFNCT FEYISDAFSL DVSEKSGNFK 180
HLREFVFKNK DGFLYVYKGY QPIDVVRDLP SGFNTLKPIF KLPLGINITN FRAILTAFSP 240
AQDTWGTSA AAYFVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY 300
QTSNFRVVPV GDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKKISNCVA DYSLVYNSTF 360
FSTFKCYGVS ATKLNLCFVS NVYADSFVVK GDDVRQIAPG QTGVIADYNY KLPDDFMGCV 420
45 LAWNTRNIDA TSTGNYNYKY RYLRHGKLRP FERDISNVPF SPDGKPCPTP ALNCYWPLND 480
YGFYTTTGIG YQPYRVVVLV FELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLTP 540
SSKRFQPFQ FGRDVSDFTD SVRDPKTSEI LDISPCSFSG VSVITPGTNA SSEVAVLYQD 600
VNCTDVSTAI HADQLTPAW IYSTGNNVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY 660
HTVSLLRSTS QKSIVAYTMS LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC 720
50 NMYICGDSTE CANLLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKDFG 780
GFNFSQLPDP PLKPTKRSFI EDLLFNKVTI ADAGFMKQYG ECLGDINARD LICAQKFNGI 840
TVLPPLLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTONVLYE 900
NQKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN 960
DILSRDLKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK 1020
55 RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVVFVN 1080
GTSWFITQRN FFSPPQIITD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN 1140
HTSPDVLGD ISGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYEY YIKWPWYVWL 1200
GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCKCFDE DDSEPVKGV KLHYT 1255

60 <212> Type : PRT

<211> Length : 1255

SequenceName : SEQ ID 274

SequenceDescription :

65 Sequence

<213> OrganismName : SARS coronavirus BJ01

<400> PreSequenceString :
MFIFLLFLTL TSGSDDLDRCT TFDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL 60
PFYSNVTGFH TINHTFDNPV IPFKDGIYFA ATEKSNVVRG WVGSTMMNK SQSVIIINNS 120
TNVIRACNF ELCDNPPFAV SKPMGTQHT MIFDNANFCT FEYISDAFSL DVSEKSGNFK 180
5 HREFVFKNK DGFLYVYKGY QPIDVVRDLP SGFNTLKPIF KLPLGINITN FRAILTAFSP 240
AQDTWGTSA AYFVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY 300
QTSNFRVVPV GDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKKISNCVA DYSLVYNSTF 360
FSTPKCYGVS ATKLNLCFS NVYADSFVVK GDDVRQIAPG QTGVIADYNY KLPDDFMGCV 420
LAWNTRNIDA TSTGNYNKY RYLRHGKLRP FERDISNVFF SPDGKPCPP ALNCYWPLND 480
10 YGFTTTGIG YQPYRVVLS FELNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLT 540
SSKRFQPFQ FGRDVSDFTD SVRDPKTSEI LDISPCSPGG VSVITPGTNA SSEVAVLYQD 600
VNCTDVSTAI HADQLTPAWR IYSTGNNVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY 660
HTVSLRSTS QKSIVAYTMS LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC 720
NMYICGDSTE CANLLLYQS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKYFG 780
15 GFNFSQILPD PLKPTKRSFI EDLLENKVTI ADAGFMKQYG ECLGDINARD LICAQKFNG 840
TVLPLLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTQNVLYE 900
NQKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN 960
DILSRDLKVE AEFVQIDRLT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK 1020
RVDFCGKGYH LMSFPQAAPH GVVLHVTVV PSQERNFTTA PAICHEGKAY FPREGVVFVN 1080
20 GTSWFITQRN FFSPQIITD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKEYFKN 1140
HTSPDVLGLD ISGINASVNV IQKEIDRLNE VAKNLNESLI DLQELGKYEY YIKWPWYVWL 1200
GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCCKFDE DDSEPVKGV KLHYT 1255

<212> Type : PRT
25 <211> Length : 1255
SequenceName : SEQ ID 275
SequenceDescription :
Sequence

30 <213> OrganismName : SARS coronavirus
<400> PreSequenceString :
SGFRKMAFPS GKVEGCMVQV TCGTTTLNGL WLDDTVYCPR HVICTAEDML NPNYEDLLIR 60
KSNHSFLVQA GNVQLRVIGH SMQNCLLRK VDTSNPKTPK YKQVRIQPGQ TFSVLACYNG 120
35 SPSGVYQCAM RPNHTIKGSF LNGSCGSVGF NIDYDCVSFC YMHHMELPTG VHAGTDLEGK 180
FYGPVDRQT AQAAGTDTTI TLNVLAWLYA AVINGDRWFL NRFTTTLNDF NLVAMKYNYE 240
PLTQDHQDIL GPLSAQTGIA VLDMCAALKE LLQNGMNGRT ILGSTILEDE FTPFDVVRQC 300
SGVTFQ 306
<212> Type : PRT
40 <211> Length : 306
SequenceName : SEQ ID 276
SequenceDescription :
Sequence

45 <213> OrganismName : SARS coronavirus
<400> PreSequenceString :
AIASEFSSLP SYAAYATAQE AYEQAVANGD SEVLKLLKK SLNVAKSEFD RDAAMQRKLE 60
KMADQAMTQM YQARSEDKR AKVTSAMQTM LFTMLRKLND DALNNIINNA RDGCVPLNII 120
50 PLTTAAKLMV VVPDYGTYN TCDGNTFTYA SALWEIQQVV DADSKIVQLS EINMDNSPNL 180
AWPLIVTALR ANSAVKLQ 198
<212> Type : PRT
<211> Length : 198
SequenceName : SEQ ID 277
SequenceDescription :
Sequence

60 <213> OrganismName : SARS coronavirus
<400> PreSequenceString :
AGNATEVPAN STVLSFCAFA VDPKAYKDY LASGGQPITN CVKMLCTHTG TGQAITVPE 60
ANMDQESFGG ASCCLYCRCH IDHPNPKGFC DLKGYVQIP TTCANDPVGF TLRNTVCTVC 120
GMWKGYGCSG DQLREPLMQ 139
<212> Type : PRT
65 <211> Length : 139
SequenceName : SEQ ID 278
SequenceDescription :

Sequence

5 <213> OrganismName : SARS coronavirus
 <400> PreSequenceString :
 NNELSPVALR QMSCAAGTTQ TACTDDNALA YYNNSKGGRF VLALLSDHQD LKWARFPKSD 60
 GTGTIYTELE PPCRFVTDTP KGPVKVLYLF IKGLNNLNRG MVLGSLAATV RLQ 113

<212> Type : PRT
 10 <211> Length : 113
 SequenceName : SEQ ID 279
 SequenceDescription :

15 Sequence

<213> OrganismName : Escherichia coli O157:H7
 <400> PreSequenceString :
 MNKIFKVIWN PATGNYTVTS ETAKSRGKKS GRSKLLISAL VAGGMLSSFG ALANAGNDNG 60
 20 QGVDYGGGSA GDGWVAIGKG AKANTFMNTS GSSTAVGYDA IAEQYSSAI GSKTHAIGGA 120
 SMAFGVSAIS EGDRSIALGA SSYSLGQYSM ALGRYSKALG KLSIAMGDSS KAEGANAIAL 180
 GNATKATEIM SIALGDTANA SKAYSMALGA SSVASEENAI AIGAETEAAE NATAIGNNAK 240
 AKGTNSMAMG FGSLADKVNT IALNGSQAL ADNIAIGQG NKADGVDAIA LGNGSQSRGL 300
 NTIALGTASN ATGDKSLALG SNSSANGINS VALGADSIAD LDNTVSVGNS SLKRKIVNVK 360
 25 NGAIKSDSYD AINGSQLYAI SDSVAKRLGG GAAVDVDDGT VTAPTYNLKN GSKNNVGAAL 420
 AVLIDENTLQW DQTKGKYSAA HGTSSPTASV ITDVADGTIS ASSKDAVNGS QLKATNDDVE 480
 ANTANIATNT SNIATNTANI ATNTNTITNL TDSVGLQAD ALLWNETKKA FSAAHGQDIT 540
 SKITNVKDDAD LTADSTDAVN GSQKKTNDNA VATNTNIN ANTSNIATNTT NISNLTETVT 600
 NLGEDALKWD KDNVFTAAH GTETTSKITN VKDGLTTGS TDAVNGSQLK TTNDAVATNT 660
 30 TNIAATNTNI SNLTETVTNL GEDALKWDKD NGVFTAAHGN NTASKITNIL DGTVTATSSD 720
 AINGSQLYDL SSNIATYFQG NASVNTDGVF TGPTYKIGET NYYNVGDALA AINSSSFSTSL 780
 GDALLWDATA GKFSAKHGTN GDASVITDVA DGEISDSSSD AVNGSQLHGV SSVVVDALGG 840
 GAENVADGTI TAPTYTIANA DYDNVGDALN AIDTTLDDAL LWDADAGENG AFSAAHGKDK 900
 TASVITNVAN GAISAASSDA INGSQLYTTN KYIADALGGD AEVNADGTIT APTYTIANAE 960
 35 YNNVGDALDA LDDNALLWDE TANGGAGAYN ASHDGKASII TNVANGSISE DSTDAVNGSQ 1020
 LNATNMIEBQ NTQIINQLAG NTDATYIQEN GAGINYVRTN DDGLAFNDAS AQGVGATAIG 1080
 YNSVAKGDSS VAIQGSYSYD VDTGIALGSS SVSSRVIAKG SRDTSITENG VVIGYDITDG 1140
 ELLGALSTGD DGKYRQIINV ADGSEAHDAV TVRQLQNAIG AVATTPTKYF HANSTEEDSL 1200
 AVGTDLSLNG AKTIVNGDKG IGIGYGAYVD ANALNGIAIG SNAQVIHVNS IAINNGSTTT 1260
 40 RGAQNTYTAY NMDAPQNSVG EFSVGSADGQ RQITNVAAQS ADTDAVNNGQ LKVTDAQVSQ 1320
 NTQSITNLND RVTNLDSRVV NIENGIGDIV TTGSTKYFKT NTDGVDASAG GKDSVAIGSG 1380
 STAAADNSVA LGTGSVATEE TTISVGSSTN RRITNVAAAG KNATDAVNVA QLKSSSEAGGV 1440
 RYDTKADGSI DYSNITLGGG NGGTTRISNV SAGVNNNDVV NYAQLKQSVQ ETKQYTDQRM 1500
 VEMDNKLSKT ESKLSGGIAS AMAMTGLPQA YTPGASMASI GGGTYNGESA VALGVSMVSA 1560
 45 NGRWVYKLQG STNSQGEYSA ALGAGIQW 1588
 <212> Type : PRT
 <211> Length : 1588
 SequenceName : SEQ ID 280
 SequenceDescription :

50 Sequence

<213> OrganismName : Escherichia coli O157:H7
 <400> PreSequenceString :
 55 MPASAVGALG EASYTVTANV TDSAGNSNSA SHNVQVNTAL PGVTINPVAT DDIINAAESG 60
 NAQTISGQVT GAAAGDTVTV TLGGKTYTAT VQGNLSWSVD VPAADIQAIG NGNLTVNASV 120
 TNGVGN TGSG SRDITIDANL PGLRVDTVAG DDVNSIEHA QALVITGSSS GLAAGAALT 180
 VINTVTYAAT VLADGTWSVG VPAADVSNWP AGTVNITVSG TNTAGTTSTI THPVTVDLAA 240
 VAISINTVSG DDVINAAEKG ADLTLGSGTS GVEVGQTVTV TFGGKTYTAT VAGDGSWTTT 300
 60 VPAADLSVLR DGDATVQASV STINGNTASA THAYSVDATA PTLAINTIAT DDILNAAEAG 360
 NPLTISGSST AEAGQTVTVT LNGVTVSGSV QADGSWSVSL PTADLSNLTA SQYTVSASVS 420
 DKAGNPASAN HGLAVDLTVP VLTINTVSGD DIINAAEHGQ ALVISGSSTG GEAGDVITVT 480
 LNSKTYTTML DASGNWSVG VPAADVTLGS GPQTITAAIT DAAGNSDDAS RTVTVNLAAP 540
 TIGINTIATD DVIKATEKGA DLQITGTSNQ PAGTTTIVTL NGQNYTATTD SNGNWSATVP 600
 65 ASAVSALGEA NYTVTANVTD TAGNSNSASH NVLVNSALPA VTINAVATDD IINAAESGNA 660
 QTISGQVGA AQGDVTVTTL GGNTYTATVQ SNLSWSVDVP AADIQALNGD DLTVNASVTN 720
 GVGNTGSGSR DITIDANLPG LRVDTVAGDD VINSIEHNQA LVITGSSSGL TAGTALTVEI 780

	NNVTYGATVL	ADGTWSLGPV	AVDVSNNWPAG	TVNITVSGTN	SAGTTSTITH	PVTVDLAGVA	840
	ITINTLSGDD	VINAVEKGET	LUVSGSTSGV	EAGQTVTVTF	GGKNYTTTVE	ANGSWTVNVV	900
	PADLAALPDG	AGNVQASVSN	INGNSAQADR	AYSVDATAPL	VTINTIASDD	ILNVSEAGAG	960
	ITISGTTTAQ	AGQTLTVTLN	NNTYQTTVLA	DGTWSVNVPA	ADLSGLTASS	YTVTATVSDK	1020
5	AGNPASADHA	LVVDITAPDL	TINTVAGDDI	INAIEHQAL	VVSGTSTGAA	AGDVVTVTLN	1080
	GKNYTTTLDA	SGNWSVGIPA	ADVTALATGS	QTITASLSDR	AGNSDSTHSD	VTYDLSGPTL	1140
	TINTVSGDDI	INAAEIVVAQ	TISGQVTGTA	VAGNTVIVTI	GGNQYNATVQ	SDLSWSVSVS	1200
	ANVLQALGNG	ELTISASLTN	SANNTGTATH	DIVIDANLPG	LRVDTVAGDD	VINSIEHTQA	1260
	LVITGSSSGL	AAGAALTUVI	NSVTYGATVL	ADGWSVSGVP	VADVTNWPAG	TVNIAVSGTN	1320
10	TAGTTTISH	PVTVDLAAVA	ITINTLSTDD	VINAEEKGSD	LQLSGTTSVG	EAGQTTITVIF	1380
	GKSYTTTVA	ADNTWGLTIP	AVDVATLPDG	AANVQASVSN	VAGNSTQATH	AYSVDATAPS	1440
	VTINTIATDD	ILNAAEAGSA	LTISGTSTAE	AGQTVTVTLN	GVNYSNGVQA	DGSWSVSVPT	1500
	GDLASLTASS	YTVNASVSDK	ARNASASATHN	LTVDLAAPVV	TINTVAGDDI	INATEHQQAQ	1560
	IISGSATGAT	TGNTVSVTIG	TTTYTTVLDA	NGNWSIGVPA	SVISALAQGD	VTITATVTD	1620
15	AGNSGTASHT	VTVALGAPVL	AINITIAVDDI	INAAEKGADL	AITGTSNQPA	GTQITVTLNG	1680
	QNYTTTADAS	GNNVSVTPAS	RVSALGEATY	TVTAAATDAD	GNSGSASHNV	QVNTALPGVT	1740
	INVVATDDII	NAAEAGVEQT	ISGQVTGAAA	GDTVTVTLGG	ATYTATVQAN	LSWSVDVPAS	1800
	ALQELGNGEL	PVTVDLAVNS	GNTGNGTREI	TIDANLPLGR	VDTVAGDDVV	NIEHGGQALV	1860
	ITGSSSGLAA	GSNVTLTING	QTYVAAVLAD	GTWSVGVPV	DVSAWPAGSV	TIAASGSTSA	1920
20	GNPVSVTHPV	TVDLASAVAS	INAITADDVI	NAAEKGAAAL	LSGSTSGVEA	GQTVTVTFGG	1980
	KTYASATVAAN	GSWSTVPA	DMAALRDGDA	SAQASVSNVN	GNSATTHAY	SVDASAPT	2040
	INTIAGDIL	NAAEAGAAAL	ITGSSSTAEAG	QTVTVTLNGT	NYTGTQVTDG	SWSVSVPSAD	2100
	LSTLTASNYT	VNAASVDKAG	NPASVNHNL	VDTSVPVVTI	NTVAGDDVIN	ATEHAQAQII	2160
	SGSATGAATG	STVTVTIGTN	TFTTVLDASG	NWSVGVPASV	VSAANGT	INASVTDAGG	2220
25	NSGSATHQVT	VNTGLPTITF	NAISGDNILN	ADEKQPLTI	SGGSTGLATG	AQVTVTLNGH	2280
	NYSATTDSAG	NWLTVPVSD	LAALGQANYT	VSAATSAA	NTASSQANLL	VDSGLPDVTI	2340
	NTVAGDDIIN	AAEAGADQTI	SGVVTRAAG	DTVTVTLGGN	TYTATVQSNL	SWSVSVPTAD	2400
	LQALGNGDIT	ITASVTNANG	NTGSGTRDIT	IDANLPLGRV	DTVAGDDIVN	SIEHGGQALVI	2460
	TGGSSGLNAG	AVLTVTINSV	AYSATVQADG	SWSVGIPAA	VSAWPAGPLT	VEVDGQSSAN	2520
30	NPVSVSHPT	VDLTAVAI	NTVASDDVIN	AAEKGTLNL	SGSTSGIESG	QTVTVTFGGK	2580
	TYTASVAANG	SWSVNVPAAD	LATLPEGAAN	VQASVSSASG	NSASATHAYS	VDASAPTLTI	2640
	NTIASDDILN	AAEAGSPLTI	SGTSTAETGQ	TVTVTLNGAT	YTGTVQADGS	WSVSVPTSAL	2700
	GALNASNTTV	SATVNDKAGN	PGSASHNLAV	DTTAPVLTIN	TVAGDDIIND	AEHAQALVIS	2760
	GTSSGGEAGD	VSVVLNKGK	YTTTLDASGN	WSVGVPAAV	TALGSGAQTI	TASVSDRAGN	2820
35	SDDASRTTV	SLAPVISIN	TIAGDDVINA	TEKSGDLALS	GTSDQPAQTA	ITVTLNGQNY	2880
	SATTDASGNW	SVTPASAVS	ALGEATYSVT	ASVINAQGNS	STASHNVQVN	TALPGITINP	2940
	VATDDIINAS	EAGSAQTISG	QVTGAAAGST	TVTLGGKTY	TATVQADLSW	NVSVPAAWDQ	3000
	ALNGELTVN	ASVTNANGNT	SGSTRDITD	ASLPLGRVDT	VAGDDVNNII	EHAQAQVITG	3060
	SSSGFAAGTA	LTVVINNQTY	AATVLANGSW	SVGVPATDVS	NWPAGTLNIT	VSGANSAGTQ	3120
40	TSITHPLTVD	LTAVAISMNS	ITSDDAINAA	EKGAAALTLG	STSGVEAGQT	VTVTFGGKTY	3180
	TTTVAANGSW	STTVPAADLA	ALRDGDASAQ	VRVTNVNGNS	ATATHEYSVD	SAAPT	3240
	IASDNIINAS	EAAAGVTVSG	TSTAQTGQTL	TVTLNGTNYQ	TTVQTDGWS	LTLPASDLTA	3300
	LANNGYTLTA	TVSDLAGNLG	SASKGVTVD	TAPVISFNTV	AGDDVINNV	HIQAQIISGT	3360
	ATGAVAGDRL	VVTIAGQQYV	TSTDASGNWS	VGVPASVISG	LADGTVTISA	TITDSAGNSS	3420
45	TQTHNVQVNT	AAVSLSVSTI	SGDNLINAAE	AGSALTLSGT	GTNFATGTVV	TVLLNGKGY	3480
	ATIQSNGSWS	VNVPAADVAA	LSDGTSYTVS	ASAQDSAGNG	NSSTQTHNVQ	VNTAAVSLSV	3540
	STISGDNLIN	AAEAGSALT	SGTGTNFATG	TVTVLLNGK	GYSATIQSNG	SWSVNVPAAD	3600
	VAALSDGTSY	TVSASAQDSA	GNSATASRSV	AVDLTAPVIS	INTVSTDDRL	NAAEQQPLT	3660
	LNGSTSAEVG	QTVTVTFGGK	TYTATVAANG	TVLNVPAVD	LAALGQGAQT	ITASVNDRAG	3720
50	NPGQATHALT	VDTVAPT	ATVAGDDIIN	NAEQLAGQTI	SGTTTAEVQ	TVTVTFNGQT	3780
	WSATVSGSGS	WSVFIPAQQF	AGLSGGSYTI	SATVSDQAGN	PGSASRGVTL	NGDVPTVTIN	3840
	TFAGDDVVNA	AEHGSSLVIS	GTTTAPVGQT	LTLTLNGKTY	TTTVQTGGSW	SYTLGSADVT	3900
	ALADGNAYVI	NASVSNAIGN	TGSSNHTITV	DLSAPAMGIN	IDSLQADTGL	SASDFITSVS	3960
	PVVVNGSLTA	ALASNETAQI	SIDGGTTWTT	LTVTGTWRY	NDSRSLTDGN	YLYQVRVIDA	4020
55	AGNVGATDSQ	NVVIDTTAPD	PAVKTIAISA	ITTDMLITN	DFVTSDTTLA	VSGTLGATLS	4080
	AGEFAQISLD	GGVTWTTLT	VGTSSWSYADG	HTLTDGTWNY	TVRVVDLAGN	VGQTATQNVV	4140
	VDTTSPAAK	SITITGISDD	TGTSSSDFIT	SDTTLTVRGV	LGAALGANEF	AQISTDNGAT	4200
	WVNVTVAA	LNWSYVDGRT	LTNGTTTWQV	RVVDLAGNVG	ATSSQSALID	TVNPAQVLT	4260
	ASISTDTGSS	ATDFITSDDT	LTLTGSLGAG	LASGEVAQIS	LDGATWTTL	TTNGTQWY	4320
60	DSRTLTGDSY	VQVQVRLDLA	GNTGPVVSKT	VVVDINPTA	TPTIVSYTDD	VGQRQGLTSS	4380
	SQATDDTTPL	LNGVLSAPLA	SGEVVYLRYN	GLLLGAVTMV	GALNWTYSDS	GLVSGAYTYS	4440
	ARVVDLAGNI	TSSSDFVLT	DTSIPTTLAQ	ITSQTRD	PIISGVITAA	LASGQYVEV	4500
	INGKTYTSEP	GGAVVVDPAH	NTWYVQLPDT	DALTVSATAY	TVTAQVKSSA	GNGNNANISN	4560
	GTVTVNAID	YTPTWTTASK	TTAWGLTYGL	DSHGMMTVLA	NQQVMQSTDP	LTWSKTALT	4620
65	YQSGNNYATS	SIADYDRNGT	GDLFITRDDY	GTGYINGFTN	NGDGTFS	QVTVGTLTWY	4680
	GSIVAFDKEG	DGYLDFWIGD	AGGPDNSFTL	WNNAGTLVGN	STSSNGGSA	TVGGAVTGYL	4740
	SLNEGSGVDL	NNDGRIDLQV	HTYNLNNYYT	LSSLINQNG	TFVWQNTTN	TFLSGAGSGA	4800

MSSSVSMTWA DFDGDGMDL FLPASQGRAN YGSLLENTNG VLGPVAVGA TATTYASQFS 4860
LAVDWNHDL MDIARIAQTG QSYLYTNVSN ASNWTQSALG GSQSGTTSKV AAMDYDWDGA 4920
VDVLVSKQSG SVFLSRNTNT VSYGTSLSHLR ITDPNGINVT YGNTVVKLYNS AGVLVATQII 4980
NPQSGMGVND TSALVNFYGL NAGETYNAVL IKSTGTASN IDQTVNTSWG GLQATDATHA 5040
5 YDLSAEAGTA SNNGKFPVGTG YNDTFFATAG TDTYDGSQGW VYSSGTGTWL ANGGMDVVDV 5100
RLSTVGV TAN LSSTAQAATG FNTSTFTNIE GISGSNFNDI LTGSSGDNQL EGRGGNDTLN 5160
IGNGGHDTLL YKLLNASDAT GGNGSDVVG FTVGTWEGTA DTDRIDIREL LQSGSGYTNG 5220
KASYVNGVAT LDAQAGNIGD FVKVTQSGSD TIVQIDRDGT GGTATTNVV TLTGVHTDLA 5280
TLLANHQMLV V 5291

10 <212> Type : PRT
<211> Length : 5291
SequenceName : SEQ ID 281
SequenceDescription :

15 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :

20 MGVTAEATL PNGNNDTKIV NIAPDASNAQ VTLNIPAAQV VTNNSDSVQL TATVKDPSNH 60
PVAGITVNT MPQDVAANFT LENNGIAITQ ANGEAHVTLK GKKAGTHTVT ATLSNNNTSD 120
SQPVTFVADK TSALVVLQIS KNEITGNGVD SATLTATVKD QFDNEVNLP VTFSTASSGL 180
TLTPGESNTN ESGIAQATLA GVAFGEQTVT ASLANNGASD NKTVHFIGDT AAKIIEIETP 240
VPDSIIAGTF QNSSGGVITA TVVDNNGFPV KGVTVNFTSN AATAEMTNGG QAVTNEQGKA 300
TVTYTNRSS IESGARPDVT EASLENGSST LSTSINVNAD ASTAHLTLLQ ALFDTVSAGD 360
25 TTNLVIEVK NYGNGVPQOE VTLVSPSEG VTPSNNAIYT TNHDGNFYAS FTATKAGVYQ 420
VTATLENGDS MQQTVTYVPN VANAEISLAA SKDPVIANN DLTLTATVA DTEGNAIANS 480
EVTFTLPEV RANFTLGDGG KVVTDTEGKA KVTLLKGTAG AHTVTASMAK GKSEQLVVNF 540
IADTLTAQVN LNVTEDNFIA NNVGMTRLQA TVTDGNGNPL ANEAVTFTLP ADVSASFTLG 600
QGGSAITDIN GKAEVTLSGT KSGTYPVTVS VNNYGVSDTK QVTLIADAGT AKLASLTSVY 660
30 SFVVTTEGA TMTASVTDAN GNPVEGIKVN FRGTSVTLSS TSVETDDRGF AEILVTSTEV 720
GLKTVSASLA DKPTEVISRL LNAKADINSA TITSLEIPEG QMVAAQDVAV KAHVNDQFGN 780
PILNESVTFE AEPPEHMTIS QNVISTDTHG IAEVTMTPER NGSYMKASL ANGSSYEKDL 840
VVLDQKLTL ASSPLIGVNS PTGATLTATL TSANGTPVEG QVINFSVTPE GATLSGGKVR 900
TNSSGQAPVV LTSNKVGTYT VTASFHNGVT IQTQITVKVT GNSSTAHVAS FIADPSTIAA 960
35 TNSDLSTLKA TVEDGSGNLI EGLTVYFALK SGSATLTSLT AVTDQNGIAT TSVRGAITGS 1020
VTVSAVTTAG GMQTVDTILV AGPADASQSV LKNRSSLKLG DFTDSAEHL VLHDISGNPI 1080
KVSEGLETFQ SGTNAPVQV SAIDYSKNFS GEYKATVTG GEGIATLIPV LNGVHQAGLS 1140
TTIQFTRAED KIMSGTVLVN GANLPTTTFP SQGFTGAYYQ LNNDNFAPGK TAADYEFSSS 1200
ASWVDV DATG KVTFTKNVGSK WERITATPKT GGPSYIYEIR VKSWVWVAGD AFMIYSLAEN 1260
40 FCSSNGYTLP LGDHLNHSRS RGIGSLYSEW GDMGHYTEA GFHSNMYWSS SPANSNEQYV 1320
VSLATGDQSV FEKLGAYAT CYKNL 1345
<212> Type : PRT
<211> Length : 1345
SequenceName : SEQ ID 282
SequenceDescription :

45 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :

50 MSLIIDVISR KTSVKQTLIN PGDVTVVIYE PSVQVHAQA SAVARYVREG NDLLIYMQDG 60
TVIRCNGYFL QAANTAEQSE LVFADGQQLT HITFADTAAG GLAPVELTAQ TTAIESIAPF 120
LDTVAQTSAP PWGWLGAAGV GGGALGALLA SGGDGSKTE VINNPTPPAE PGNATPSFLV 180
TDNQGDQGI LATNDITDDT TPTFSGSGQA GATIQIKDSN GNTIASTQVD NNGHWSVSLP 240
55 TQSAGEHTWS VVQIVGSTIT DAGSITLTID NSQASQVAT TAGDNIINAS EQAAGFTLSG 300
TSSHLAQGTE LTVTLNGKTY TTSVGANGAW SVQVPTADAQ ALGEGNQAVL VSGKDATGNT 360
VTGAQLLTVD TQPTLAINT IAQDNIIASA EHNVALVLSG TSNAEAGQTV TLTVNGKSHT 420
ATVSGDGTWQ VTLPADEVQA LAEGNYAVNA SVSDRAGNTT SHSANFTVDT SAPVVSNTV 480
AGDDILNNAE QAVAQIISGQ VSGASPGDTV TVKLGTHTVLT GIVLADGSWN VALDPAVTRT 540
60 LDRGANTIFV TVTDAAGNTG AASRAITLVG VSPLITINTV SGDDIISGAE KGAPLTLTG 600
TQQAETGQTV TVTLAQGSFT TTVQADGSWS LTVPAAAMGN LPPDGAIVTA SVTDLSGNTG 660
NTSRITIVDS QAPALSIDPL TADNIIAAE SGQDLPTGT TDAQPGQTV TLTNGQTYQG 720
VVQPDGTWSV TVPAANVGAL ADGNATVTAS VNDVAGNPSS VSRVALVDAT PPVVTINPVA 780
TDNVINTPEH AQAQIISGTV TGAQAGDIVT VTLNNVDYTT VVDGSGNWSL GVPASVVSGL 840
65 ADGSPVSVS VTDKAGNTGS QSLTIVTNT APLIGINIA GDDVINASEK GADLQITGTS 900
DQPVNTAITV TDLNGQYTTT TDASGNWSVT VPASAVTALG QANYTVTAAV TSDIGNSATA 960
SHNVLVDSAL PGVTINPVAT DDIINAAEAG VAQTISGQVT GAEDGDTVTT TLGGNTYTAT 1020

VGSNLTWSVD VPAADIQALG NGDLTVNASV TNQNGNTGSG TRDITIDANL PGLRVDTVAG 1080
DDVUNIIEHG QALVVVTGSSS GLAESTPLTV TINNVEYTTA VQADGSWSVG VTAAQVSAWP 1140
AGTVNIAVSG ESSAGNSVSI THPVTVDLTP AAITINTIAT DDVINAEEKG ADLTLSTGTTT 1200
NVEPGQTVTV TFGGKNYAS VASDGSWTAT VPAADLASLP EGSASALASV SNINGNSASA 1260
5 VHNYSVDSSA PTIIINTVAS DNIVNASEAD AGVTVSGSTT AEAGQIVTIT LNSPTVQTYQ 1320
ATVQADGSWS INIPAADLEA LTDGSHTLTA TVNDKAGNPA STTHNLAVDL TVPVLTTINTI 1380
AGDDIINATE HGQALVISGS STGGEAGDVV TVTLNSKTYT TTLDASGNWS VGVPAADVTA 1440
LGSGPQTVTA TVTDAAGNSD N 1461
<212> Type : PRT
10 <211> Length : 1461
SequenceName : SEQ ID 283
SequenceDescription :

Sequence

15 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MNRIYRVWNL CTLQVFQACS ELTRRVGKTS TVNLRKSSGL TTKFSRLTLG VLLALSGSVS 60
GASLEVDNGQ ITNIDTDVAY DAYLVGWYGT GVLNLAGGN ASLTITTSV IGGNEDSEGT 120
20 VNVLGGTWRL YDSGNNARPL NVGQSGTGTL NIKQKGHVDG GYLRLGTQAA GVGTNVVEGE 180
DSVLTTLEFE IGSYGTGSLN ITDKGYVTSS IVAILGYQAN SNGKVVEKG GEWLKNNDS 240
SIEFQIGNQG TGEATIREGG LITAENTIIG GNATGVGTLN VQDQDSVITV RRLYNGYFGN 300
GAVNISNNGI INNKEYSLVG VQDGSBGVVN VTDKGHWNFL GTGEAFRIY IGDAGDGELN 360
VSREGKVDSD IITAGMKETG TGNLTVKDKN SVITNLGTNL GYDGHGEMNI SNEGLVVSNG 420
25 GSSLGYGETG VGKVSITGG IWEVNKNVYT TIGVAGVGNL NISDGGKFVS QNITFLGDKA 480
SGIGTLNLMG ATSSFDTVGI NVGNFGSGIV NVSNGATLNS TGYGFIGGNA SGKGIVNIST 540
DSLWNLKTSS TNAQLLQVGV LGTGELNITT GGIVKARDTQ IALNDKSKGD VRVDGQNSLL 600
ETFNMYVGTG GTGTTLTLNS GTLNVEGGEV YLGVFEPAVG TLNIGAAHGE AAADAGFITN 660
ATKVEFGSGE GVVFVNHTNN SDAGYQVDM LITGDDKDGKV IHDAGHTVFN AGNTYSGKTL 720
30 VNDGLLTIAS HTADGVTGMG SSEVTIASPG TLDILASTNS AGDYTLTNAL KGDGLMRVQL 780
SSSDKMFQFT HATGTEFAGV AQLKDSFTTL ERDNTAALTH AMLQSDIENT TSVNVGEQSI 840
GGLAMNGGTL IFDITDIPAA LAEGYISVDI LTVGASDVTW KGRNYQVNGT GDVLIGVPKP 900
WNDPMANNPL TTLNLLEHDD NHVGVQLVKA QTVIGSGGSL TLRDLQGDEV EADKTLHIAQ 960
NGTVVAEGDY GFRLTTPAGD GLYVNYGLKA LNIHGGQKLT LAEHGGAYGA TADMSAKIGG 1020
35 EGDLAINTVR QVSLNQGQND YQATYVQMG TLRTDADGAL GNTRELNINIS AAIVDLNGST 1080
QTVEFTTGQM GSTVLFKEGS LTVNKGGISQ GELTGGGNLN VTGGTLAVEG LNARYNALTS 1140
VSPNAEVSLE NTQGLGRGNI ANDGLLTLKN VTGELRNSIS GKGIVSATAR TDVELDGDNS 1200
RFVQGQFNIDT GSALSVNEQK NLGDASVINN GLLTISTERS WAMTHSISGS GDLTKLGTGI 1260
LTLNNDSSAY QGTTDIVGGE IAFGSDSAIN TASQHINIHN SGVMSGNVTT AGDVNVMSGG 1320
40 TLRVAKTTIG ESAATWRMAA RFK 1343
<212> Type : PRT
<211> Length : 1343
SequenceName : SEQ ID 284
SequenceDescription :

Sequence

45 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
50 MGIKQHNNGT KADRLAELKI RSPSIQLIKF GAIGLNAILF SPLLIAADTG SQYGTNITIN 60
DGDRITGDTA DPSGNLYGVM TPAGNTPGNI NLGNDVTVNV NDASGYAKGI IIQKKNSSLT 120
ANRLTVDVVG QTSAGINLI GDYTHADLGT GSTIKSNDDG IIIGHSSTLT ATQFTIENSN 180
GIGLTINDYG TSVDLGSQSK IKTDGSTGVY IGGLNGNNAN GAARFTATDL TIDVQGYSAM 240
GINVQKNSVV DLGTNSSIKT SGDHAHGLWS FGQVSANALT VDVTAAGANG VEVRGGTTTI 300
55 GADSHISSAQ GGGLVTSQSD ATINFSGTAA QRNSIFSGGS YGASAQTATA VINMQNTDIT 360
VDRNGSLALG LWALSGGRIT GDSLAIITGAA GARGIYAMTN SQIDLTSGLV IDMSTPDQMA 420
IATQHDDGYA ASRINASGRM LINGSVLSKG GLINLDMHPG SVWTGSSLSL NVNGGKLDA 480
MNNSVWNVTS NSNLDLALS HSTVDFASHG STAGTFTTLN VENLSGNSTF IMRADVVGE 540
NGVKPWA 547
60 <212> Type : PRT
<211> Length : 547
SequenceName : SEQ ID 285
SequenceDescription :

Sequence

65 <213> OrganismName : Escherichia coli O157:H7

<400> PreSequenceString :
MGIDSRNDIP EGIATLGAFM GYSHSHIGFD RGGHGSVDSY SLGGYASWEH ESGFYLDGVV 60
KLNRFESNVA GKMSGGGAAN GSYHSNGLGG HIETGMRFTD GNWNLTPTYAS LTGFTADNPE 120
YHLSNGMESK SVDTRSIYRE LGATLSYNMR LGNGMEVEPW LKAAVRKEFV DDNRVKVNSD 180
5 GNFBVNDLSGR RGIYQAGIKA SFSSTLSGHL GVGYSNGAGM ESPWNAVAGV NWSF 234

<212> Type : PRT
<211> Length : 234
SequenceName : SEQ ID 286
10 SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
15 <400> PreSequenceString :
MKKKVLALAL VTVFTGMGVA QAADVTAQAV ATWSATAKKD TTSKLVVTPL GSLAFQYAE 60
IKGFNSQKGL FDVAIEGDST ATAFKLTSRL ITNTLTQLDT SGSTLNVGVD YNGAAVEKTG 120
DTVMIDTANG VLGGNLSPLA NGYNASNRTT AQDGFTFSII SGTNGTTAV TDYSTLPEGI 180
WSGDVSVQFD ATWTS 195
20 <212> Type : PRT
<211> Length : 195
SequenceName : SEQ ID 287
SequenceDescription :

25 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MTAESYDDNY LDDEDADWTA TGQGQKSAGD TSFTLAWKPG EEGQKGLIGW FESGDVRA 60
30 IRFPNGTVDV FRGWVSSIGK AVTAKEVITR TVKVTNVGKP SVAEERSKIT PVSAIKVTPT 120
SGTVAKGKTT TLTVSFEPES ATDKTFRAVS ADPSKATISV KDMTITVNGV ATGKVQIPVV 180
SGNGQFAAVA EVTVTEAGAA G 201
<212> Type : PRT
<211> Length : 201
35 SequenceName : SEQ ID 288
SequenceDescription :

Sequence

40 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MTAESYDDNY LDDEDADWTA TGQGQKSAGD TSFTLAWKPG EEGQKGLIGW FESGDVRA 60
IRFPNGTVDV FRGWVSSIGK AVTAKEVITR TVKVTNVGKP SVAEERSKIT PVSAIKVTPT 120
SGTVAKGKTT TLTVSFEPES ATDKTFRAVS ADPSKATISV KDMTITVNGV ATGKVQIPVV 180
45 SGNGQFAAVA EVTVTEAGAA G 201
<212> Type : PRT
<211> Length : 201
SequenceName : SEQ ID 289
SequenceDescription :

50 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
55 MLYNIPCRIY ILSTLSLCIS GIVSTATATS SETKISNEET LVVTNRSAS NLWESPATIQ 60
VIDQQTQNS TNASIADNLQ DIPGVEITDN SLAGRKQIRI RGEASSRVLI LIDGQEVTYQ 120
RAGDNYGVGL LIDESALERV EVVKGPSVL YGSAIGGIV NFITKKGGDK LASGVVKAVY 180
NSATAGWEES IAVQGSIGGF DYRINGSYSD QGNRDPDGR LPNTNYRNNS QGVWLGYN 240
NHRFGLSLDR YRLATQTYYE DPDGSYEAFS VKIPKLEREK VGVFYD TDVD GDYLLK 300
60 AYEQTIQRQF ANEVKTTQPV PSPMIQALTV HNKTDT HDKQ YTQAVTLQSH FSLPANN 360
TGAQYKQDRV SQRSGGMTSS KSLTGFINKE TRRSYYESE QSTVSLFAQN DWQFADH 420
TMGVRQYWLS SKLTRGDGVS YTAGIISDTS LARESASDHE MVTSTSLRYS GFDNLE 480
FAQGVFPTL SQLFMQTSAG GSVTYGNPDL KAEHSNNFEL GARYNGNQWL IDSAVY 540
KDYIASLICD GSIVCNGNTN SSRSSYYYD NIDRAKTWGL EISAEYNGWV FSPYISGN 600
65 RRQYETSTLK TTNTEGPAIN GRIGLKHTLV MGQANIISDV FIRAASSAKD DSNGTET 660
GWATLNFAVN TEFGNEDQYR INLALNNLTD KRYRTAHETI PAAGFNAAIG FVWNF 715

<212> Type : PRT
<211> Length : 715
SequenceName : SEQ ID 290
SequenceDescription :

5
Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
10 MTKMSRYALI TALAMFLAGC VGQREPAPVE EVKPAPEQPA EPQQPVPTVP SVPTIPQQPG 60
PIEHEDQTAP PAPHIRHYDW NGAMQPMVSK MLGADGVTAG SVLLVDSVNN RTNGSLNAAE 120
ATETLRNALA NNGKFTLVSA QQLSMAKQQL GLSPQDSLGT RSKAIGIARN VGAHYVLYSS 180
ASGNVNAPTL QMQLMLVQTG EIIWSGKGAV SQQ 213
<212> Type : PRT
15 <211> Length : 213
SequenceName : SEQ ID 291
SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
20 MKSKVLALLI PALLGAGAAH AAEVYNKDG N KLDLYGKVDG LHYFSDNSAK DGDQSYARLG 60
FKGETQINDQ LTGYGQWEYN IQANNTSSK NQSWTRLAFA GLKFS DYGSF DYGRNYGLDR 120
YAA 123
<212> Type : PRT
<211> Length : 123
SequenceName : SEQ ID 292
SequenceDescription :

30
Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
35 MATPNPLEPV KGAGTTLWVY NGKGDAYANP LSDDDWQRLA KVKDLTPGEM TAEPYDDNYL 60
DDEDADWTAT GQKQKSAGDT SFTLAWKPGE EGQKGLIGWF ESGDV RAYKI RFPNGTV DVF 120
RGWVSSIGKA VTAKEVITRT VKVTNVGKPS VAEERSEITP ATAIKVTPTS GTVAKGKTTT 180
LTVSFEPESA TDKTFRAVSA DPSKATISVK DMTITVNGVA TGVQI PVVS GNGQFAAAVE 240
VTVTEAGAAG 250
40 <212> Type : PRT
<211> Length : 250
SequenceName : SEQ ID 293
SequenceDescription :

45
Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
50 MATPNPLEPV KGAGTTLWVY NGKGDAYANP LSDDDWQRLA KVKDLTPGEM TAEPYDDNYL 60
DDEDADWTAT GQKQKSAGDT SFTLAWKPGE EGQKGLIGWF ESGDV RAYKI RFPNGTV DVF 120
RGWVSSIGKA VTAKEVITRT VKVTNVGKPS VAEERSEITP ATAIKVTPTS GTVAKGKTTT 180
LTVSFEPESA TDKTFRAVSA DPSKATISVK DMTITVNGVA TGVQI PVVS GNGQFAAAVE 240
VTVTEAGAAG 250
<212> Type : PRT
55 <211> Length : 250
SequenceName : SEQ ID 294
SequenceDescription :

60
Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
65 MGWTDMLPEF GGSYTNADN FMTGRANGVA TYRNTDFGFL VNGLNFAVQY QGNNEGASNG 60
QEGTNNGRDV RHENGDWGL STTYDLGMGF SAGAA YTSSD RTNDQVNH TA AGGDKADAWT 120
AGLKYDANNI YLATMYSETR NMTPFGSDY AVANKTQNF E VTAQYQFDFG LRPVAVSFLMS 180
KGRDLHAAGG ADNPA GVDDK DLVKYADVGA TYYFNKNMST YVDYKINLLD EDDSFYAANG 240
ISTDDIVALG L VYQF 255

<212> Type : PRT
<211> Length : 255
SequenceName : SEQ ID 295
SequenceDescription :

5
Sequence

<213> OrganismName : Haemophilus influenzae Rd
<400> PreSequenceString :

10 MGFIMKLT KT ALCTALFATF TFSANAQTYP DLPVGIKGGT GALIGDTVYV GLGSGGDKFY 60
TLDLKDPSAQ WKEIATFP GG ERNQPVAAAV DGKLYVFGGL QKNEKGELQL VNDAYRYNPS 120
DNTWMKLPTR SPRGLVGSSG ASHGDKVYIL GGSNLSIFNG FFQDTVAAGE DKAKKDEIAA 180
AYFDQRPEDY FFTTELLSYE PSTNKWRNEG RIPFSGRAGA AFTIQGNELV VVNGEIKPGL 240
RTAETHQKGF TAKGVQWKNL PDLPAKPKGS QDGLAGALSG YSNGHVLTG GANFPKSIKQ 300
15 FKEGKLHAK GLSKAWHNEV YTLNNGKWRI VGELPMNIGY GFSVSYNNKV LLIGGETDGG 360
KALTSVKAIS YDGKKLTIE 379
<212> Type : PRT
<211> Length : 379
SequenceName : SEQ ID 296
SequenceDescription :

20
Sequence

<213> OrganismName : Haemophilus influenzae Rd
<400> PreSequenceString :

25 MGEQYMLTTI LSLIVTTTV AYVSWLKTG DDLKSSKGYF LAGRGLSGLV IGCSMVLTSI 60
STEQLIGVNA VSYKGNFSVI AWTVPTVIPL CFLALYIIGW L 101
<212> Type : PRT
<211> Length : 101
SequenceName : SEQ ID 297
SequenceDescription :

30
Sequence

<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :

35 MKNQHKNPLT KALMKTYPYN HFLFFCFILG AFLLGLLSPA YALSIITTKE IDANLLNGAI 60
ESRVVLGKRV FKVEAHGFYF RNNATNSIDI EITSLLRDNQ SFPLTSSAKT SLKIPPNNAKI 120
KKSTILVLKG ENAEVAKIL GVSKEEYQKL ENIAQTKAAN DPMYANTPFS NGSDSSFYDN 180
40 NPNSPNNAI NGKDGANGSN GYGANGNDGV NGISGSNGAN GSHSNNAIG SGIDTDGVLG 240
VDGVNGSSSS SGGSVGGYEN NFTNHGSTNN NTGGYDNFNN GSSSGGSLGN GGLFPPIPGN 300
GDTNNSNNST NTTSPNGSS SNNATNPSSQ ENNYSSQYCK VPELSPNNTM KLDVIAKDGS 360
CISMNALRDD TKCAYRYDFE AGKAIKQTQY YYVDRENKTQ NIGGCVDLQG AQYAMQLYKD 420
DSKCALQTTS DKG YGMGKTQ TFQTEIVFRG MDNLIHVAVP CSDYARVQDR IVRYEKNDKT 480
45 QTLTPIVDQY YNDPNNPNKQ EILNRGIATQ LSSQYQEFAC QQWEYNDAKL EAKRPTMLKS 540
YNKLNGEWVE VTPCNFEAGI KSGAVVSPYV MGVPSSKVLS DITTSHYFRI ERKNYGEREQ 600
CQKLYGVNRC QPQYSILILV SPIGAPLTKP LPPKPLNLIY AQP KIMKNTP QPIILSPLKP 660
PSTGLKAF 668
<212> Type : PRT
<211> Length : 668
SequenceName : SEQ ID 298
SequenceDescription :

50
Sequence

<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :

55 MPVIRVLVLM ATMMMLVKVT AKEKKVFKNV GISIMGIAFW EAIKDSIKKQ IKKSDWICGN 60
VKTADDYLKT HPNSWFNSAI GVTAITAMLM NVCFADDQSK KEVAQAQKEA ENARDRANKS 120
60 GIELEQEEQK TEQEKQKTEQ EKQKTEQEKQ KTEQEKQKTE QEKQKTSNIE TNNQIKVEQE 180
QQKTEQEKQK TNNQKDLVN KAEQNCQENH NQFFIKKLG I KAGIAIEIEA ECKTPKPTKT 240
NQTPIQPKHL PNSKQPHSQ RSKAQELIAY LQKELESPLP SQKAIKQVD FYRPSIAYL 300
ELDPRDFNAT BEWQKENLKI RSKAQAKMLE MRS LKPDPOA HLSTSQSLLL VQKIFADVSK 360
EIKVVANTEK KVEKAGYGYS KRM 383
<212> Type : PRT
<211> Length : 383
SequenceName : SEQ ID 299

SequenceDescription :

Sequence

5 <213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MNYPNLPNSA LEISEQPEVK EITNELLKQL QNALRSNAHF SEQVELSLKC IVRILEVLLS 60
LDFFKNANEI DSSLRNSIEW LTNAGESLKL KMKEYERFFS EFNTSMHANE QEVNTNLNAN 120
AENIKSEIKK LENQLIETTT RLLTSYQIFL NQARDNANNQ ITKNKTQSLE AITQAKNNAN 180
10 NEISNNQTQA ITNITEAKTN ANNEISNNQT QAITNINEAK ESATTQINAN KQEAINNITQ 240
EKTQATSEIT EAKKTDHYQN IDFFEFE 267
<212> Type : PRT
<211> Length : 267
SequenceName : SEQ ID 300
15 SequenceDescription :

Sequence

20 <213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKFFSKDLFK KVTPLFLSVY FLSPTLTQAK SRFYVASQYQ VGKMIMKKYN DLKRTIEGAS 60
FSLGWEINPT NYWFYSRYYP FMDYGNVILN KRTGAQANMF TYGFGGDLIM EYNKNPLYVF 120
SLFYGMQVAE NTWTISKHSA NFIIDDWRSI QGFSLKTSNF RMLGLVGFKF QTVLFHHDAS 180
IEVGIKWPFA FEYDSPFVRL FSVFISHTFY L 211
25 <212> Type : PRT
<211> Length : 211
SequenceName : SEQ ID 301
SequenceDescription :

Sequence

30 <213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKKFTLSLFL CCTLNAAED IFRNNTNETD LTNSFEHGKE NNNLIPAKSD SLESFKEQEN 60
35 KEKAKQIMDL KALQSVYFSK NRKLQDNNFN VLYVAGNTNK IRLRYAMTTT FIFDNDPIIY 120
VSLGDPDSFE LTYPTNDHYD LSNMLVIKPL LIGVDTNLTV VGASGTIYTL LFV 173
<212> Type : PRT
<211> Length : 173
40 SequenceName : SEQ ID 302
SequenceDescription :

Sequence

45 <213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MLDYVPWIGN GYRYGNNHRG SNSSTSGVTT QGQSQNASSN EPAPTFSNVG VGLKANVNGT 60
LSGSRTTFNQ QGTPWLTLTQ ANLQLWTGAG WRNDKNGQSD ENYTNFASAK GSTNQGGSTT 120
GGSAGNPDSL KQDKADKSGD SVTVAEATSG DNLNTYTNLP PTSPPHPTDR TRCHSPTRTT 180
50 PSGCSCSCAA CWAASRCWSI RVGKMIVTSL IPPTKNGLTP N 221
<212> Type : PRT
<211> Length : 221
SequenceName : SEQ ID 303
SequenceDescription :

Sequence

55 <213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
60 MDDITAPQTS AGSSSGTSTN TSGSRSFLPT FSNVGVGLKA NVQGTLLGGRQ TTTTGNNIPK 60
WATLDQANLQ LWTGAGWRND KTTSGSTGNA NDTKFTSATG SGSGQGSSSG TSTSAGNPDG 120
LQADKVDQNG QVKTQVEAT SGDNLNTYTN LPPANLTPTA DWPNALSFNT KNNQAQQLF 180
LRGLLGSIPIV LVNKGQDDN SKFKAEDQKW SYTDLQSDQT KLNLPAYGEV NGLLNPAIVE 240
TYFGNTRASG SGNTTSSPG IGFKIPEQSG TTTTSKAVLI TPGLAWTPQD VGNIVVSGTS 300
65 FSFQLGGWLV TFTDFIKPRA GYLGLQLTGL DVSEATQREL IWAKRPWAAF RGSWVNRLLR 360
VESVWDFKGV WADQAQLAAQ AATSTTTTA TGATLPEHPN ALAYQISYTD KDSYKASTQG 420
SGQTNSQNNS PYLHFIKPKK VESTTQLDQG LKNLLDPNQV RTKLRSFSGT DHSTQPPQPS 480

LKTTTPVFGR SSGNLSSVFS GGGAGGGSSG SGQSGVDLSP VERVSGH 527
<212> Type : PRT
<211> Length : 527
SequenceName : SEQ ID 304
SequenceDescription :

5 Sequence

<213> OrganismName : Mycoplasma pneumoniae
10 <400> PreSequenceString :
MLKLAVGIFI SPTLTRFSTG FNLAGSVLDQ VLDYVPWIGN GHRYGNHRG VDDITAPKTG 60
AGSSSGTSTN TSGSRSLPT FSNVGVGLKA NVQGTLLGGSQ TTTTGKDIPK WPTLDPANLQ 120
LWTGAGWRND KASNKQSDEN HTTFKSATGS GQGGSTTGG SAGNPDSLKQ DKISKSGQNL 180
TTQDGAQSN STTESASND HLPNLTPTS DWPNALSTN KNNQAQALF LRGLLGSIPV 240
15 LVNRSGSDDS NKFQATDQKW SYTDLKSDQT KLNLPAYGEV NGLNLPALVE TYFGTTRAGG 300
SGSNTTSSPG IGFKIPEQNN DSKAVLITPG LAWTPQDVGN LVVSGTSLSF QLGGLVLTFT 360
DFVKPRAGYL GLQLTGLDAS DATQRALIWA KRPWAAFRGS WVNRLGRVES VWDLKGWQD 420
QAQAAAQAAAT TAAATGDALP EHPNALAYQI SSTDKDSYKA STQSSGQTN QNTSPYLHLI 480
KPKKVENTTQ LDQGLKTCWT PTRFAPSCAK ALVQTIPPKP NPNPSKQPHR CLGRIVVTLA 540
20 VCLVVGVL EE QTAPIRWTSP PLNGWVGGLW GNYFVGVGGI VVRILKVCKT LFFISIFISI 600
FFLNCSLTFL IWTASLATG LTVVGHFTST TTTLKRQQFS YTRPDEVALR HTNAINPRLT 660
PWTYRNTSFS SLPLTGENPG AWALVRDNTA KGITAGSGSQ QTTYDPTRTE AALTATTTFV 720
LRRYDLAGRC TTSTFRS 737
<212> Type : PRT
25 <211> Length : 737
SequenceName : SEQ ID 305
SequenceDescription :

30 Sequence

<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MLDYIPWIGN GHRYGNDHRG SNSSTSGVTT QGQQSQNASG TEPASTFSNV GVGLKANVQG 60
TLGGSQTTTT GKDIPKWPTL DQANLQLWTG AGWRNDKASS QSDENHTKF TSATGSGQQG 120
35 SSSGTTSNAG NPDSLKQDKV DKSGDSVTVA ETTSGDNLTN YTNLPPNLT TADWPNALSF 180
TNKNNQAQRAQ LFLRALLGSI PVLVNKSGQD DSNKFQATDQ KWSYTELKSD QTKNLNPAYG 240
EVNGLLNPAL VEVYGLSSTQ GSSTGAGGAG GNTGGDTNTQ TYARPGIGFK LPSTDSESSK 300
ATLITPGLAW TAQDVGNLVV SGTSLSFQLG GWLVTFTDFI KPRSGYLGLO LTGLDANDSD 360
QRELIWAPPA LNRSLWQLGQ PLGPRGECVG FQGGVGGSSS VRLASSYKYH HRNEGYLIGA 420
40 HQCFGLSGEL YRPGFVQGFH SKLRPKPKHL PLPALGAGEK SRFLW 465
<212> Type : PRT
<211> Length : 465
SequenceName : SEQ ID 306
SequenceDescription :

45 Sequence

<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
50 MLGSIPVLVN RSGSDSNKFQ ATDQKWSYTD LQSDQTKLNL SAYGEVNGLL NPALVETYFG 60
TTRTSSTANQ NSTVPGIGF KIPEQNNSDK ATLITPGLAW TPQDVGNLVV SGTTVSFQLG 120
GWLVTFTDFV KPRAGYLGLO LSGLNASDSD QRELIWAPRP WAAFRGSWVN RLGRVESVWD 180
LKGWVADQAQ LAAQAATSSST TTTATGATLP EHPNALAYQI SYTDKDSYKA STQSSGQTN 240
QNNSLYLHLI KPKKVESTTQ LDQGLKNLLD PNQVRTKLRLQ SFGTDHSTQP QPQSLKTTTP 300
55 VFGAMSGNLG SVLSGGGAGG AGSTNSVDLS PVERVSGSLT INRNFSY 347
<212> Type : PRT
<211> Length : 347
SequenceName : SEQ ID 307
SequenceDescription :

60 Sequence

<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
65 MGQGGQSGTS AGNPDSLKQD KISKSGSLT TDQGNATGQQ EATNYTNLPP NLTPADWPN 60
ALSFTNKNNA HRAQLFLRGL LGSIPVLVNR SGSDSNKFQA TDQKWSYTDL QSDQTKLNL 120
AYGEVNGLLN PALVETYFGN TRAGGSGSNT TSSPGIGFKI PEQNNSKAT LITPGLAWTP 180

QDVGNLVVSQ TSLSFQLGGW LVSFTDFIKP RAGYLGLQLS GLDASDSQQR ELIWAKRPWA 240
 AFRGSWVNLRL GRVESVWDLK GVVADQAQLA AQAATSEASG SALAPHPNAL AFQVSVVEAS 300
 AYSSSTSSSG SGSSSNTSPY LHLIKPKKVE STTQLDQGLK NLLDPNQVRT KLRQSFGTDH 360
 STQPQSLTT TPVFGTSSGN IGSVLSSGGA GGGSSSGSQS GVDLSPVERV SGH 413

5 <212> Type : PRT
 <211> Length : 413
 SequenceName : SEQ ID 308
 SequenceDescription :

10 Sequence

 <213> OrganismName : Mycoplasma pneumoniae
 <400> PreSequenceString :

15 MGLQLSGLDA SDSDQRELIW AKRPWAAFRG SWVNRLGRVE SVWDLKGVWA DQAHSAVSES 60
 QAATSSSTTT ATGDTLPEHP NALAYQISST DKDSYKASTQ GSGQTNSQNT SPYLHLIKPK 120
 KVTASDKLDD DLKNLLDPNE VRVKLRQSFQ TDHSTQPPQ PLKTTTPVFG TNSGNLGSVL 180
 SGGGTTQDSS TTNQLSPVQR VSGWLVGQLP STSDGNTSST NNLAPOINTG NEVVGVDLS 240
 KRASIESRL WIALKP 256

20 <212> Type : PRT
 <211> Length : 256
 SequenceName : SEQ ID 309
 SequenceDescription :

25 Sequence

 <213> OrganismName : Mycoplasma pneumoniae
 <400> PreSequenceString :

30 MRDNTAKGIT AGSGSQTTY DPARTEATLT TTFALRRYD LAGRALYDL FSKLNPQTPT 60
 RDANCQITFN PFGGFGLSGS APQQWNEVK KVPVEVAQDP TDPYRFAVLL VPRSVVYEQ 120
 LQRGLALPNQ GSSSGSQQN TTIGAYGLKV KNAEADTAKS NEKLQDESK SSGSSSTST 180
 TTQRGSTNSD TKVKALKIEV KKKSDSEDNG QLQLEKNDLA NAPIKRGEES GQSVQLKADD 240
 FGTAPSSSGS GGNSNPGSPT PWRPWLATEQ IHKDLPKWSA SILILVDAPY ARNRTAIDRV 300
 DHLDPKVMTA NYPPSWRMPK WNHHLWDWK ARDVLFTQTG FDESNTSNTK QGFQKEADSD 360
 35 KSAPIALPFE AYFANTGNLT WFGQALLVFG GNGHVTKSAH TAPLSIWLYI YLVKAVTFRL 420
 LLANSLLSKS NIYKKTAN 438

<212> Type : PRT
 <211> Length : 438
 SequenceName : SEQ ID 310
 SequenceDescription :

40 Sequence

 <213> OrganismName : Mycoplasma pneumoniae
 <400> PreSequenceString :

45 MRDNIAGKIT AGSNTQTTY DPTREATLT TATTFALRRY DLAGRALYDL DFSKLNQPTP 60
 TRDQTGQITF NPFGGFGLSG AAPQWNEVK DKVPVEVAQD PSNPYRFAVL LVPRSVVYEQ 120
 QLQRGLALPN QGSSSGSQQN NTIGAYGLK VKNAEADTAK SNEKLQYES KSSNGSSSTS 180
 TTQRGGSSNE NKVKALQVAV KKKSGSQGNS GDQTEQVEL ESNDLANAPI KRGSNNNQV 240
 50 QLKADDFGTA PSSSGSGTQD GTPTPWPWL TTEQIHNDPA KFAASILILY DAPYARNRTA 300
 IDRVDHLDPK VMTANYPPSW RTPKWNHGL WDWKARDVLL QTTGFFNPRR HPEWFDGGQT 360
 VADNEKTGFD VDSENTKQG FQKEADSDKS APIALPFEAY FANIGNLTWF EQALLVFGIC 420
 LS 422

<212> Type : PRT
 55 <211> Length : 422
 SequenceName : SEQ ID 311
 SequenceDescription :

60 Sequence

 <213> OrganismName : Mycoplasma pneumoniae
 <400> PreSequenceString :

65 MLWPFWRVWW KRVLTQTRA PAKPNPLTVP PTCTWWSLRK LPNPTKLDDD LKNLLDPNEV 60
 RARMLKSFGT ENFTQPPQP QALKTTTPVF GTSSGNLGSV LSGGGYHAGL KHHQSTVTRS 120
 TGEWVDR 127

<212> Type : PRT
 <211> Length : 127

SequenceName : SEQ ID 312
SequenceDescription :

5 Sequence

<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MRDNSAKGIT AGSESQTTY DPTRTEAALT ASTTFALRRY DLAGRALYDL DFSRLNPQTP 60
TRDQTGQITF NPFGGFGLSG AAPQQWNEVK NKVPVEVAQD PSNPYRFAVL LVPRSVVYYE 120
10 QLQRGLALPN QGSSSGSGQQ NTTIGAYGLK VKNAEADTAK SNEKLQGES KSSNGSSSTS 180
TTTQRGGSSG DTKVKALQVA VKKKSGSQGN SGEQGTEQVE LESNDLANAP IKRGEESGQS 240
VQLKAADFQT TPSSSGSGGN SNPGSPTPWR PWLATEQIHK DLPKWSASIL ILYDAPYARN 300
RTAIDRVDDL DPKVMTANYP PSWRTPKWNH HGLWDWKARD VLLQTTGFFN SRRHPEWFDQ 360
GQAVADNTQT GFDTDDTDNK KTRLKSGSWL RQAGPDRPPV WSVLRQHWQP HTVRASAFGV 420
15 WDLFVLIN 428

<212> Type : PRT
<211> Length : 428
SequenceName : SEQ ID 313
SequenceDescription :

20 Sequence

<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
25 MFGLKVKNAE ADTAKSNEKL QGAETGSST TSGSGQSTQR GGSSGDTKVK ALQVAVKKKS 60
GSQNGSGDQG TEQVELESND LANAPIKRGs NPASPTQGSR LRHHPIQFGI WSIRHPHPLK 120
AVACDRANSQ GPPQMIRLDP HSVRCALCL 149

<212> Type : PRT
<211> Length : 149
SequenceName : SEQ ID 314
SequenceDescription :

Sequence

35 <213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MFGLKVKDAT VDSSKQSTES LKGESSSSSS TTSSTSTTQR GGSSGDTKVK ALQVAVKKKS 60
DSEDNGQIEL ETNNLANAPI KRGSNNNQV QLKADDFGTS PSSSESGQSG TPTPWTPLA 120
TEQIHKDLPK WSASILILYD APYARNRTAI DRVDHLDPKV MTANYPPSWR TPKWNHHGLW 180
40 DWKARDVLVQ TTGFFNPRRH PDWFDQGQAV AENTQTGFDT DDTDNKKQGF RKQGEQSPAP 240
IALPFEAYFA NIGNLTWFGQ ALLVFGICLS 270

<212> Type : PRT
<211> Length : 270
SequenceName : SEQ ID 315
SequenceDescription :

Sequence

50 <213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MGSQNQGSTT TTSAGNPSL VTDKVDQKQ VQTSQNLSL TNYTNLSPNF TPTSDWPNAL 60
SFTNKNNQAR AQLFLHGLLG SIPVLVKNKG ENNEKFQATD QKWSYTELKS DQTKLNLPA 120
GEVNGLLNPA LVETYFGTTR TSSTANQNST TVPGIGFKIP EQNNDKAVL ITPGLAWTPQ 180
DVGNLVVSGT SFSFQLGGWL VSFTDFVKPR AGYLGLQLTG LDASDATQRA LIWAPPALSG 240
55 LSWQLGQPVG PRGECVGFEG GVGSSSVRL ARIYHHRNRG YLTGAPECFG LSGECGSEC 300
LQAKHELPRN PIH 313

<212> Type : PRT
<211> Length : 313
SequenceName : SEQ ID 316
SequenceDescription :

Sequence

65 <213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MSFGLVGTVN NNGWKSPFRH ETKYRAGYDK FKYYKTHYRG AKKAGTNDNR WRWTAWFDLD 60
FAHQKIVLIE RGELHRQADL KKSDPATNET SKTVWGSIKE KLLQNVNHLH SEKGVFLWFR 120

QSGFTTTRN
<212> Type : PRT
<211> Length : 129
SequenceName : SEQ ID 317
SequenceDescription :

5 Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
10 MAEPLAVDPT GLSAAAAKLA GLVFPQPPAP IAVSGTDSVV AAINETMPPI ESLVSDGLPG 60
VKAALTRTAS NMNAAADVYA KTDQSLGTSL SQYAFGSSGE GLAGVASVGG QPSQATQLLS 120
TPVSQVTTQL GETAAELAPR VVATVPQLVQ LAPHAVQMSQ NASPIAQTIS QTAQQAQSA 180
QGGSGPMPAQ LASAEKPATE QAEPPVHEVTN DDQGDQGDVQ PAEVVAAARD QGAGGSGGQQ 240
15 PGGGVPAQAM DTGAGARPAA SPLAAPVDPS TPAPSTTTTL 280
<212> Type : PRT
<211> Length : 280
SequenceName : SEQ ID 318
SequenceDescription :

20 Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
25 MRYLIATAVL VAVVLVGWPA AGAPPSCAGL GGTVQAGQIC HVHASGPKYM LDMTFPVDYP 60
DQQALTDYIT QNRDGFVNVA QGSPLRDQPY QMDATSEQHS SGQPPQATRS VVLKFFQDLG 120
GAHPSTWYKA FNYNLATSQP ITFDTLFPVQ TTPLDISIYPI VQRELARQTG FGAAILPSTG 180
LDBAHYQNFA ITDDSLIFYF AQGELLPSFV GACQAQVPRS AIPPLAI 227
<212> Type : PRT
<211> Length : 227
SequenceName : SEQ ID 319
SequenceDescription :

30 Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
35 MKMVKSIAAG LTAAAAIGAA AAGVTSIMAG GPVVYQMOPV VFGAPLEPLDP ASAPDVPTAA 60
QLTSLNLNSLA DPNVSFANKG SLVEGGIGGT EARIADHKLK KAAEHGDLPL SFSVTNIQPA 120
40 AAGSATADVS VSGPKLSSPV TQNVTFVNQG GWMLSRASAM ELLQAAGN 168
<212> Type : PRT
<211> Length : 168
SequenceName : SEQ ID 320
SequenceDescription :

45 Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
50 MTYSPGNPGY PQAQAGSYG GVTPSFAHAD EGASKLPMYL NIAVAVLGLA AYFASFGPMF 60
TLSTELGGGD GAVSGDTGLP VGVALLAALL AGVALVPKAK SHVTTVAVLG VLGVFLMVSA 120
TFNKPSAYST GWALWVVLAF IVFQAVAAVL ALLVETGAIT APAPRPKFDP YGQYGRYGQY 180
GOYGVQPGGY YGQQAQQAQ GLQSPGPQQS PQPPGYGSQY GGYSSSPSQS GSGYTAQPPA 240
QPPAQSGSQQ SHQGPSTPPT GFPSFSPPPP VSAGTGSQAG SAPVNYSNPS GGEQSSSPGG 300
55 APV 303
<212> Type : PRT
<211> Length : 303
SequenceName : SEQ ID 321
SequenceDescription :

60 Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
65 MKCPGVSDCV ATVRHDNVFA IAAGLRWSAA VPPLHKGDV TKLLVGAIAG GMLACAAILG 60
DGIASADTAL IVPGTAPSPY GPLRSLYHFN PAMQPQIGAN YYNPTATRHV VSYPGSFWPV 120
TGLNSPTVGS SVSAGTNNLD AAIRSTDGPI FVAGLSQGTI VLDREQARLA NDPTAPPPGQ 180

LTFIKAGDPN NLLWRAFRPG THVPIIDYTV PAPAESQYDT INIVGQYDIF SDPPNRPGNL 240
LADLNAIAAG GYYGHSATAF SDPARVAPRD TTTTNSLGA TTTYFYRTD QLPLVRALVD 300
MAGLPPQAAG TVDAALRPPII DRAYQPGPAP AVNPRDLVQG IRGIPAIAPA IAIPIGSTTG 360
ASAATSTAAA TAAATNALRG ANVGPGANKA LSMVRGLLPK GKKH 404

5 <212> Type : PRT
<211> Length : 404
SequenceName : SEQ ID 322
SequenceDescription :

10 Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSTGCTTAAAG GPPFFDAIPN FIEDLDVIVA AAPPAPACST GVSAAQLGEI TSTTAAALQQ 60
15 KAPHCPAESD QTPAGAAGDG DLPEVGGRVT SPPQPPVAAL TGYSANIGGL SVPHSWNLPP 120
AVRQVAAMFP GATPMYMTGS SDGSYAGLAA AGLAGTGLAG LAARGGSAPT PAAAAAPAGAG 180
GAGPAATRPA AQTTPAVPAA AAGSAIPGLP PGLPPGVVAN LAATLAAIPG ATIIVVPPSP 240
NANQ 244
<212> Type : PRT
20 <211> Length : 244
SequenceName : SEQ ID 323
SequenceDescription :

25 Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MDVALGVAVT DRVARLALVD SAAPGTVIDQ FVLDVAEHPV EVLTETVVGT DRSLAGENHR 60
LVATRLCWPD QAKADELQHA LQDSGVHDVA VISEAQAATA LVGAAHAGSA VLLVGDETAT 120
30 LSVVGDEPDAP PTMVAAPVA GADATSTVDT LMARLGDQAL APGDVFLVGR SAEHTTVLAD 180
QLRAASTMRV QTPDDPTFAL ARGAMAAGA ATMAHPALVA DATTSLPRAE AGQSGSEGEQ 240
LAYSQASDYE LLPVDEYEEH DEYGAAADRS APLSRRSLLI GNAVVAFAVI GFASLAVAVA 300
VTIRPTAASK PVEGHQNAQP GKFMPLLPQT QQAPVPPPPP DDPTAGFQGG TIPAVQNVVP 360
RPGTSPGVGG TPASPAPEAP AVPGVVPAPV PIPVPIIIPP FPGWQPGMPT IPTAPPITPV 420
35 TTSATTPPTT PPTTPVTTTP TTPPTTPVTT PPTTPPTTPV TTPPTTVAPT TVAPTIVAPT 480
TVAPTIVAPA TATPTTVAPQ PTQQFTQQPT QQMTTQQQTV APQTVAFAAQ PPSGGRNGSG 548
GGDLFGGF 548
<212> Type : PRT
<211> Length : 548
40 SequenceName : SEQ ID 324
SequenceDescription :

Sequence

45 <213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MKNARTLLIA AAIAGTLVTT SPAGIANADD AGLDPNAAAG PDAVGFDPNL PPAPDAAPVD 60
TPPAPEDAGF DPNLPPPLAP DFLSPPAEEA PPVPVAYSVN WDAIAQCESG GNWSINTGNG 120
50 YGGLRFTAG TWRANGSGS AANASREEQI RVAENVLRSQ GIRAWPVCGR RG 172
<212> Type : PRT
<211> Length : 172
SequenceName : SEQ ID 325
SequenceDescription :

55 Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
60 MTRLIPCTL VGLMLTLLPA PTSAAGSNTA TTLFPVDEV T QLEHTTFLDC HPNGSCDFVA 60
GANLRTPDGP TGFPPLWAR QTTEIRSTNR LAYLDAHATS QFERVMKAGG SDVITTVYFG 120
EGPPDKYQTT GVIDSTNWST GQPMTDVNVV VCTHMVQVYP GVNLTSPSTC AQANFS 176
<212> Type : PRT
65 <211> Length : 176
SequenceName : SEQ ID 326
SequenceDescription :

Sequence

5 <213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MTPGLLTTAG AGRPRDRCAR IVCTVFIETA VVATMFVALL GLSTISSKAD DIDWDAIAQC 60
ESGGNWAANT GNGLYGGLQI SQATWDSNGG VGSPAAASPQ QQIEVADNIM KTQGPRAWPK 120
CSSCSQGDAP LGSLLHILTF LAAETGGCSG SRDD 154
10 <212> Type : PRT
<211> Length : 154
SequenceName : SEQ ID 327
SequenceDescription :

Sequence

15 <213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MMQQAIVSGIT GALGGAVGGV MGELTQLPQQ AMQAGQGAMQ PLMSALQQTY GAEGLDVADG 60
ARLVDSIEGE PGLGGEPEGAG DVGAGGGGGG TTPTGYLGPP PVPTSSPPTT PAGAPAKSVT 120
20 PDPVSGTPRA SGPAGMTGMP MVPPGALGAG AEGANKDKPV EKRVTGCAEW STGQGPIINST 180
AECSGEICRR QAGGHQVDAT DPCCAERRQG 210
<212> Type : PRT
<211> Length : 210
SequenceName : SEQ ID 328
25 SequenceDescription :

Sequence

30 <213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MIRELVTTAA ITGAAIGGAP VAGADPQRYD GDVPGMNYDA SLGAPCSSWE RFIFGRGPSG 60
QAEACHFPPP NQFPPAETGY WVVISYPLYGV QQVGAPCPKP QAAAQSPDGL PMLCLGARGW 120
QPGWFTGAGF FPPEP 135
<212> Type : PRT
35 <211> Length : 135
SequenceName : SEQ ID 329
SequenceDescription :

Sequence

40 <213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MKTTGTTIKL GIVWLVLVSF TVMLIIVVFGQ VRFHHTTGYS AVFTHVSGLR AGQFVRAAGV 60
EVGKVAKVTI IDGDKQVLVD FTVDRLSLSD QATTASIRYL NLIGDRYLEL GRGHSGQRLA 120
45 PGATIPLEHT HPALDLALL GGFRLPLFQTL DPKVNSIAS SIITVFQGGG ATINDILDQT 180
ASLTATLADR DHAIGEVVNN LNTVLATTVK HQTEFDRTVD KLEVLITGLK NRADPLAAAA 240
AHISSAAGTL ADLLGRIVHC CTAASGTSRA SSSRS 275
<212> Type : PRT
<211> Length : 275
50 SequenceName : SEQ ID 330
SequenceDescription :

Sequence

55 <213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MTPRSLVRIV GVVVATTLAL VSAPAGGRAA HADPCSDIAV VFARGTHQAS GLGDVGEAFV 60
DSLTSQVGGG SIGVYAVNYP ASDDYRASAS NGSDDASAHI QRTVASCPNT RIVLGGYSQG 120
ATVIDLSTSA MPPAVADHVA AVALFGEPSG GFSSMLWGGG SLPTIGPLYS SKTINLCAPD 180
60 DPICTGGGNI MAHVSYVQSG MTSQAATFAA NRLDHAG 217
<212> Type : PRT
<211> Length : 217
SequenceName : SEQ ID 331
SequenceDescription :

65

Sequence

<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MISTTRIDFL WILSVAFASM IALATLLTLI NQVVGTPYIP GGDSPAGTDC SELASWVSNA 60
ATARPVFGDR FNTGNBEEAAL AARGFQQGTA PNALVIGWNG HHTAVTLPDG TPVSSGEGGG 120
5 VRVGGGGAYQ PKFTHHMYLP MDVDAGEDQP PAPDEPVTAV DDVEPEMPAP CPTQRPPVTP 180
RHNLCNKLRT MPGALSAALA AAAPVWPAPI SGRGFSTSL LAKRNHPVIV GK 232

<212> Type : PRT
<211> Length : 232
10 SequenceName : SEQ ID 332
SequenceDescription :

Sequence

15 <213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MTTMITLRRR FAVAVAGVAT AAATTVTTLAP APANAADVYG AIAYSGNGSW GRSWDYPTRA 60
AAEATAVKSC GYSDCKVLTS FTACGAVAN DRAYQGGVGP TLAAAMKDAL TKLGGGYIDT 120
WACN 124
20 <212> Type : PRT
<211> Length : 124
SequenceName : SEQ ID 333
SequenceDescription :

25 Sequence

30 <213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MAGLNIYVRR WRTALHATVS ALIVAILGLA ITPVASAATA RATLSVTSTW QTGFIAFTI 60
TNSSTAPLTD WKLEFDLPAG ESVLHTWNST VARSGTHYVL SPANWNRIIA PGGSATGGLR 120
GGLTGSYSPP SSCLLNGQYP CT 142
<212> Type : PRT
<211> Length : 142
SequenceName : SEQ ID 334
35 SequenceDescription :

Sequence

40 <213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MLTRAIKTQL VLLTVLAVIA VVVLGWYFLR IPSLVGIGRY TLYAELPRSG GLYRTANVTY 60
RGITIGKVTG VEPTERGARA TMSIDNGYQI PTDASANVHS VSAVGEQFVD LVSTRTSGPY 120
LRHGQTITTT TVPSQIGPAL DAANRGLAVL PKDRVASVLH EASEAVGGIG SSLNRLIEAT 180
QAIADHVRGS LEDIDDIER SAPIIDSQVN SCNEIARWAA NLNTLAAQTA QTDPAVRSL 240
45 ANAAPTADQV NATFSDVRES LPQTLANLEV VIDMLKRYHN GVEQALVFLP QSGAIAQSVT 300
TEFFPGQAGLG VGGLALNQPP PCLTGFLPAS EWRSPADTST APLPKGTYCR IPMDASNVVR 360
GARNNPCVDV PGKRAATPRE CRSNEAYVPG GTNPWYGDEN QMLSCPAPAA RCDQPVKPGQ 420
VIPAPSVNNG INPLPADQLP GTPPPVNDPL QRPGSGTVQC NGQQPNPCVY TPSTFPTTIY 480
DVQSGKVVP DGVVYSVEAS THAGADGWKV MLAPTG 516
50 <212> Type : PRT
<211> Length : 516
SequenceName : SEQ ID 335
SequenceDescription :

55 Sequence

60 <213> OrganismName : Rickettsia prowazekii
<400> PreSequenceString :
MLNNTQFLNL MKSYMKEPEFY MSSIKNTTNL DLSSITNTIQ KAMNIFFTTN KISTESMQSL 60
FKKNSEIION NINTILNSTK EVINSKDFKQ ATEYHQKCVK SIYETSMDNA KELANIAEYA 120
SNKIFEAAANK HITKNIHNAS NNIHNTAEQV QKNFNNKSA 159
<212> Type : PRT
<211> Length : 159
SequenceName : SEQ ID 336
65 SequenceDescription :

Sequence

<213> OrganismName : Rickettsia prowazekii
<400> PreSequenceString :
5 MNIKLVITYFL ILVSSLKVNA DLNHIQDSFK YQAEQLTIE LPWNDCTAIH KFLEEKLFPS 60
EQQIKKENKI HEKYKQFYLO HNNKLSDFSM QFLEKKSEIN SVETLISGFL KFCEDNFQTS 120
KSKSHSLNFF QKQQDQWLHN IRNENYKTTY KKKYEDNTFR NIN 163
<212> Type : PRT
<211> Length : 163
SequenceName : SEQ ID 337
10 SequenceDescription :

Sequence

<213> OrganismName : Rickettsia prowazekii
15 <400> PreSequenceString :
MKKLLLIATA SATILSSSVS FAECIDNEWY LRADAGVAMF NKEQDKATGV KLKSNKAIP 60
DLGIGYYISE NVRADLTIGT TIGGKLKKYG AATNTHFTGT NVSVSHKPTV TRLLINGYVD 120
LTSFDMFDVF VGGGVGPALV KEKISGVSGL ASNTKNKTNF SYKLIFGTSA QIADGVKVEL 180
AYSWINDGKT KTHNVMYKGA SVQTGGMRYQ SHNLTVGVRP GI 222
20 <212> Type : PRT
<211> Length : 222
SequenceName : SEQ ID 338
SequenceDescription :

25 Sequence

<213> OrganismName : Rickettsia prowazekii
<400> PreSequenceString :
30 MKKNMRKQML KIISIIISL LLSSCSESTR DENGLLTDSQ STIIRDYIIS QNSKNLKVNL 60
KEKFGSNLKG VKLIGIKLTN EDLSGIDFTS CEILRTDFMG SNLEKAILTN SVIQESNFAD 120
SVIKNISGYN ADFQGSIFNN ITLQNTNFVQ SNFSDTAFNK STIINVNFEN SKFSNVLWCH 180
SNIDSSNFQK THLKNNSFKN TNVMNSIFYG ADLGKSVINN TNFTNNYFES SDLSNTKFTS 240
VLIKDSNFTQ SIFNSVNFNN IQSNNSFFSY TSFEDSTLHN IHLTKCDLQN STINSSVFNN 300
FKIDNAILTN MSLNDNTFNN LSIKNSNTNF VRINKSKGFN ITLLNTNYSN AIFSNNDLKE 360
35 PKVINTDLNN SEIINSNFTN GQFNNVNFSS SLIQNVNFTD VKITLGNLQ VALINSNLIN 420
TNIINSVLBN SQINNINYQA YYSFINTNVS NNIVINDNSN QIPPNIVIN SEKDLQNISN 480
LANMNLTFN LSNLVPNGVD FSKSIFKKAN LTNVVIKNSI LKDNFSAAI LTKTDFSKSI 540
LTGSIFKFAQ IDQTCFSNSD LTNTDFTEAT IKNTAFDNAN THGIGGLE 588
40 <212> Type : PRT
<211> Length : 588
SequenceName : SEQ ID 339
SequenceDescription :

Sequence

45 <213> OrganismName : Porphyromonas gingivalis W83
<400> PreSequenceString :
MIQKFTNVKL NDMRKILSFL MMCSLHLGLQ SQTWHGDPDS VAALPSIGIQ ESSCTRITFE 60
VVFPGFYSVE KREGNQVFQR ISMPGCGSFG NLGEAELPVL KKMIAVPEFS TANVAVKIKE 120
50 TETFDNYNIY PNPTYVVEEL PEGGTYLVEA FAINNDYYSQ NVSLPSTHYV YSQDGYFRSQ 180
RFIEVTLYPF RYNPVRQEIL FAKKIEVTIT FDNPPPLQK NTGIFNKVAS SAFINYEADG 240
KSAIENDMVF SRGTTYISG NVASNLPQNC DYLVIYDDMF NVNQPHDEI KRLCEHRAFY 300
NGFDVAAVSI KDVLNSFPSN ATSYINETKL KNFIRSVYNQ SNAKRTLDGK LGYVLLIGKP 360
LSKYLADTDN TKVPTFSIHN VSLIPSHPTF GSICADSYFF SCVSPLDTVG DLFGRFSVT 420
55 NAHELHNLIE KTINKAISYN PIAHKNIYA EKGKCDAPIL RLFLKEIASG YTVNSILKSN 480
QVSAIDSIFD CLNNGSHHFP FNTGGMPTVW GIGQGLDVNT LTARLNTTSS QGLCTSLSCS 540
SAVADSTIRS LGEVLTTYAP NKGFSAPLGG SRATQYAVYL EGPCPPSEFY EYLPYSLYHN 600
LSTVVGEMLL SSIINTNSVD TYSKFNFNLL GDPALNIMAH GMEVSNCTIL PNNTIISPI 660
TIKNGGCLKI PEKGVLFHTN NGSIQVMSGG TLEIGNQAKI SGETGANPTF ITVYGDGLAI 720
60 NKQVEIDNID RLNLFTHSV MPKPHFDSVK FNSAPLYTTN CIVEISNCEF TNRSDIISK 780
CDLSVENSMP SSSGITVFKP MATSSITGLS TKAKITDNTF FATGNFAYHI TNTPGLTATS 840
NAAIKLDNIP EYVISGNKIV NCDEALVLNN SGNRTNRLHN ITRNVIKNCR IGSTLYNSYG 900
IYNRNKISNN HIGVRLNNS CFYFDNAPVI NEEDKQTFIS NRTWQLYSSN GTFPLNFHYN 960
SLQGGDTDTW IYNDTYTNR IYDVSNNHWGN NDLFDPNQVF NTPDLFIWIP FWDGLPNGRS 1020
65 GNSSAEAVEF QTALDCIGNS DYLSAKVALK MMVETYPESD FAIAALKELF RIEKMSGNDY 1080
BGLKDYFRSN PTIISQNLF PTADFLSARC DIVCENYQSA IDWYENRLNS EISYQDSVFA 1140
VIDLGDYWN MQLDSLRTGT IDLNILSCEQ RKSLESHQNV KNYLLSTLPE STGTLLPPL 1200

CNKSSLDKSK IISISPNAK AVVTIIYYTD NPSCSVIKIY GINGASADIT GLPKHLSEGY 1260
YSIQFNTSNF DPGFYLVTLN VDQKIIDTEK LRIK 1294
<212> Type : PRT
<211> Length : 1294
5 SequenceName : SEQ ID 340
 SequenceDescription :

Sequence

10 <213> OrganismName : Shigella flexneri 2a str. 2457T
 <400> PreSequenceString :
MQGKNTIVTT GDYSIGLLSQ TSGNLNTDTI IRVNSDGSVT PSFSDGDDTF IVTAGNHA VG 60
VLACASPGSA CACVSSLDEE STADTGSNEN NAIAKLDMAK GEITTHGTES YAAYANGTVV 120
KAGDTLDYTN ASVTLDVDI TTHGDNHAHAI AARQGTVSFN QGEITYTTGPD AAIKIYNGG 180
15 TVTLKNTSAV AHQSGGIVLE SSINGQEATV DILSGSSLRS ANEILYHKDE TSNVTITDSE 240
VSSAADVFIN NIKGHLTVDA TNSKITGSAN ISTDDNTHTY LSLSDNSTWD IKADSTVSNL 300
TVDNSTVYIS RADGRDVEPT RLITITENYVG NNGVLHLRTE LDDDNSATDK VVINGNTSGT 360
TRVKVTNAGG SGAYTLNGIE IISVEGESNG EFIKDSRIFA GAYEYSLTRG NTEATNKNWY 420
LTNFQATSGG ETNSGGSSAP TVAPTPVLRP EAGSYVANLA AANTLFVMRL NDRAGETRYI 480
20 DPVTEQERSS RLWLRQIGGH NAWRDSNGQL RTTSHRYVSQ LGGDLITGGF TSDSWRLGV 540
MAGYARDYNL THSSVSDYRS KGSVRGYSAG LYATWFADDI SKKGAYIDSW AQYSWFKNSV 600
KGDELAYESY SAKGATVSLE AGYGFALNKS FGLEAAKYTW IFQPPAQAIW MGVDHNAHTE 660
ANGSRIENDA NNNIQTRLGF RTFIRTQEK N SGPHGDDFEP FVEMNWIHNS KDFAVSMNGV 720
KVEQDGVSNL GEIKLVNGN LNPAASVWGN VGVQLGDNGY NDTAVMVGLK YKF 773
25 <212> Type : PRT
 <211> Length : 773
 SequenceName : SEQ ID 341
 SequenceDescription :
30 Sequence

 <213> OrganismName : Shigella flexneri 2a str. 2457T
 <400> PreSequenceString :
35 MTKLKLLALG VLIATSAGVA HAEGKFSLGA GVGVEHPYK DYDTDVYPVP VINYEGDNFW 60
FRGLGGGYL WNDATDKLSI TAYWSPLYFK AKDSGDHQMRL HLDRKSTMM AGLSYAHFTQ 120
YGYLRITTLG DTLDNSNGIV WDMAWLYRYT NGGLTVP TGI GVQWNSNQN EYYYGVSRKE 180
SARSGLRGYN PNDWSPPYLE LSASYNFLGD WSVYGTARYT RLSDEVTDSP MVDKSWTGLI 240
STGITYKF 248
40 <212> Type : PRT
 <211> Length : 248
 SequenceName : SEQ ID 342
 SequenceDescription :
45 Sequence

 <213> OrganismName : Shigella flexneri 2a str. 2457T
 <400> PreSequenceString :
50 MKKIALAGLA GMLLVASVN AMSISGQAGK EYTNIGVGFG TESTGLALSG NWTNDDDDGD 60
VAGVGLGLNL PLGPLMATVG GKGVYTNPNY GDEGYAAAVG GGLQWKIGNS FRLFGEYYS 120
PDSLSSGIQS YEEANAGARY TIMRPVSIEA GRYLNLSGK DGNRDNAVAD GLYVGVNASF 180
 <212> Type : PRT
 <211> Length : 180
55 SequenceName : SEQ ID 343
 SequenceDescription :
60 Sequence

 <213> OrganismName : Shigella flexneri 2a str. 2457T
 <400> PreSequenceString :
65 MTTLTARVFT TAEIIYRKTV IALVCHLNCS RQETVTMKNK IMALAIMMAS FAANASVLPE 60
TPVPFKSGTG AIDNDTVYIG LGSAGTAWYK LDTQAKDKKW TALAAPGGP REQATSAFID 120
GNLYVFGGIT KNSEGLTQVF NDVHKYNPKT NSWVKLSHA PMGMAGHVT F VHNKAYVTG 180
GVNQNIFNGY FEDLNEAGKD STAIIDKINAH YFDKKAEDYF FNKFLLSFDP STQQWSYAGE 240
SPWYGTAGAA VVNKGDKTWL INGEAKPGLR TDAVFELDF T GNNLKWNKLD PVSSPDGVAG 300
GFAGISNDSL IFAGGAGFKG SRENYQNGKN YAHEGLKKS Y STDIHLWHNG KWDKSGELS Q 360

GRAYGVSLPW NNSLLIIGGE TAGGKAVTDS VLISVKDNKV TVQN 404
<212> Type : PRT
<211> Length : 404
SequenceName : SEQ ID 344
5 SequenceDescription :

Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
10 <400> PreSequenceString :
MATGGAALAG KAVMGAAAGA AGGASALQAA FQKASASMET GDMSSMGSV VSSGGNGGGE 60
AGTAGSSPFA QAAGFGDSGS SSSGGGFAKA AKLATGTASE LAKGVGSQVK QGFQERVSET 120
TGGKLAASIR ESMEPKASQ SGQFEGNSLG ADSGPDSENEV RS 162
<212> Type : PRT
15 <211> Length : 162
SequenceName : SEQ ID 345
SequenceDescription :

Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MKRVLIPGVI LCGADVAAV DDKNMYMYFF EEMTVYAPVP VPVNGNTHYT SESIERLPTG 60
NGNISDLLRT NPAVRMDSTQ STSLNQGDRI PEKISIHGAS PYQNAYLIDG ISATNNLNPA 120
25 NESDASSATN ISGMSQGYL DVSLLDNVTL YDSFVPVEFG RFNGGVIDAK IKRFNADDSK 180
VKLGVRTTRL DWLTSIDEN KNSAFNQGS GSTYFSPDFK KNFYTLNFSN ELADNFGVTA 240
GLSRRQSDIT RADYVSNDSI VAGRAQYKNV IDTALSFTW FASDRFTHDL TLKYTGSSRD 300
YNTSTFPQSD REMGNKSYGL AWDMDTQLAW AKLRRTTVGWD HISDYTRHDH DIWYTELSCT 360
YGDITGRCTR GGLGHISQAV DNYTFKTRLD WQKFAVGDSV HQPYFGAEYI YSDAWTERHN 420
30 QSESYVINAA GKKTNHTIYH KGKGLGIDN YTLYMADHIS WRNVSLMPGV RYDYNLYSN 480
HNISPRFMTI WDIFADQISM ITAGYNRYYG GNILDMGLRD IRNSWTESVS GNKTLTRYQN 540
LKTPYNDELA MGLQKIDKN VIARASEAHD QISKSSRTDS ATKTTITEYN NDGKTKTHSF 600
NLSFELAEPL HIRQVDINPQ IVFSYIKSKG NLSLNNGYEE SNTGDNQVVY NGNLVSYDSV 660
PVADFNNPLK ISLNMDFTHQ PSGLVWANTL AWQEARARI ILGKTNAQYI SEYSDYKQYV 720
35 DEKLDSSLTW DTRLNTPQF LKQQLNTISA DILNVLDSTK AVDTNTTGVA TYASGRFTWL 780
DVSMKF 786
<212> Type : PRT
<211> Length : 786
SequenceName : SEQ ID 346
40 SequenceDescription :

Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
45 <400> PreSequenceString :
MKKTLALIML AGTAFASQAG TLVSQGTAS ANLTLTKPIV VNNTIQPVKG VYSGTLTAWT 60
PLATGIVGAS DGQSHDYAVT FPDDIYAESS TSADAVISGD NNPDKHLKVS LTTLEQDPPS 120
AASEEIGGKR YMMLKNTGTG GAYRVVSHMK EQVVEPDSYT IRTQAYIYAE 170
<212> Type : PRT
50 <211> Length : 170
SequenceName : SEQ ID 347
SequenceDescription :

Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
55 <400> PreSequenceString :
MGIYHWSRKT KMKRTKSIRH ASFRKNWSAR HLTVALAVA TVFMLAGCEK SDETVSPLYQN 60
ADDCSAANPG KSAECTTAYN NALKEAERTA PKYATREDCV AEFGEQGCQQ APAQAGMAPE 120
60 NQAQAQSSG SFWMPLMAGY MMGRMLGGGA GFAQQPLFSS KNPASPAYGK YTDATGKNYG 180
AAQPGRTMTV PKTAMAPKPA TTTTVTRGGF GESVAKQSTM QRSATGTSSR SMGG 234
<212> Type : PRT
<211> Length : 234
65 SequenceName : SEQ ID 348
SequenceDescription :

Sequence

<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
5 MTKMSRYALI TALAMFLAGC VGQREPAPVE EVKPAPEQPA EPQQPVPTVP SVPTIPQQPG 60
PIEHEDRTAP PAPHIRHYDW NGAMQPMVSK MLGADGVTAG SVLLVDSVNN RTNGSLNAAE 120
ATETLRNALA NNGKFTLVSA QQLSMAKQQL GLSPQDSLGT RSKAIGIARN VGAHYVLYSC 180
ASGNVNAPTL QMQLMLVQTG ETIWSGKGA V SQQ 213
<212> Type : PRT
10 <211> Length : 213
SequenceName : SEQ ID 349
SequenceDescription :

Sequence

15 <213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MTKLMQFVQR CYYMTNKKMY FILILVFTLL QVCFFALWKA RDGSTTSLEC TSTLTRNAKT 60
DHSLYYSANL SVILKKDGS SFTIVGLTDE DTPRKFSHSY FFTYKIDSNG RISGNAKAKV 120
20 SGTENQIKDE NFRNLNPLDAS LTGKGNARLS KFNNVYIFSI PGLIINTCAP I 171
<212> Type : PRT
<211> Length : 171
SequenceName : SEQ ID 350
25 SequenceDescription :

Sequence

30 <213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MGRISSGGMM FKAITTVAAAL VIATSAMAQD DLTISLAKG ETTKA AFNQM VQGHKLPAWV 60
MKGGTYTPAQ TVTLGDETYQ VMSACKPHDC GSQRIAMWS EKSNNQMTGLF SAIDEKTSQE 120
KLTWLNVNDA LSIDGKTVLF AALTGSLENH PDGFNFK 157
<212> Type : PRT
35 <211> Length : 157
SequenceName : SEQ ID 351
SequenceDescription :

Sequence

40 <213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MKKQFLEKAV FTVAATAATV VLGKMKADAD TYTLQEGDSF FSVAQRYHMD AYELASMNGK 60
DITSLILPGQ TLTVNGSAAP DNQAAAPTDT TQATTETNDA NANTYPVGQC TWGVKAVATW 120
45 AGDWWGNGGD WASSASAQGY TVGNTPAVGS IMCWTDDGGY HVAYVTAVGE DGKVQVLESN 180
YKDQQWVDNY RGWFDPNNSG TPGSVSYIYP N 211
<212> Type : PRT
<211> Length : 211
SequenceName : SEQ ID 352
50 SequenceDescription :

Sequence

55 <213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MSIKNILENK TTTIKVSFAG IATAASLILP MAVQAETTYT VKSGDTLSEI ASTHGTTVDK 60
LAKLNKINNI HLIHAGQILE LDAATEDTDA TPVQESQINE AETSASAKTS QTSEVTTTAP 120
VQESQTSEVI TSAPAETSQT SEVPTEANQT NEVSSAVSVE TSQTSEATTS APVETSQTSE 180
ATTAEPTETK TSQTNEVAAS AEENQTTSTNT SGLSTSDAAA KEFIAQKESG GNYNAKNGQY 240
60 YGRYQLSDSY LNGDLSEENQ ERVADAYVSS RYGSWTAAQA FWNANGWY 288
<212> Type : PRT
<211> Length : 288
SequenceName : SEQ ID 353
SequenceDescription :
65

Sequence

<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MKCQAFEDFK ATSLNKLSTY TGGATDGEII ANRMLQKAT KGEITMYTWN IIQNGWVNSL 60
VSWGIGGYNS SIGYSAQGNR GFSNYPYDVS MDSDNSSSSS NTGGYVNYN QSFNSGW 117

5
<212> Type : PRT
<211> Length : 117
SequenceName : SEQ ID 354
SequenceDescription :

10
Sequence

<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
15 MRYSQICKRS LALLATGMIL TTSTLPSISI LAEDSTGAPA RPDGQAPAGG GANTTTYDYS 60
GINSGLVAN GSKVTSSSKT KSTTSAQNTA LVQNGGSLTL HKANLIKSGD DNNGDNDNFY 120
GINSILLAVN ERSKAYVNSN KLKASSSGSN GIFATDKATI YANKTSIATT ADNSRGLDAT 180
YNGNIIANKM AISTKGAHSA AIATDRGGGN ISTTNSSLNT SGSGSPLLYS TGNIQVNHVT 240
GTSSNSQIAG MEGLNTILIH NSNLISTMTN KTASDPIANG VIIYQSQSGD AEATTGQSAH 300
20 FELSKSKLTS SITSGSMFYI TNSANIILN QSTLNFDANK AKLLTVAGNS ANNWGTPGSN 360
GATVNFTGCHK QTLKGDVDVD SISTLNMYLL DKTNYTGKTA VSTNSTNIS STSPITMNIS 420
KNSKWVLTGH STVTNLNAEK GAKIVDKDGK TVSVISSSQG KLVKGKSKYS LTVTGTYSQK 480
VTTSSSNKPS SSIYINRSDFD NYFKTTTAFV NNTKNTSN 518
<212> Type : PRT
25 <211> Length : 518
SequenceName : SEQ ID 355
SequenceDescription :

Sequence

30 <213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MNKIGDTLRD ARIEKLSTFD DVVDKGTGIAP HYILAMELDQ LKLLPEGKTN EYLEKYAHAV 60
GLDPVSIHIG YRNQEMSDEL ILPSSAELAA SSDSNIEKKN EGKSIEEPQE LAIDSLDVTQ 120
35 NITEETPQIE DFKVESEES KKEIKIPSRK SKYDYDEEPK KKFPWALILL ILLALTIISY 180
VGYYVYNQLQ TDSNKTLEST STKRSKDTKN DANSTTQSQT SITTFADGQ NNITLSNTNG 240
KVEVTFITLG DEESWVSATN TTDGESGTTT TATDKTYTVT LAEGSTTSM LTVGSPSGVEI 300
TINGQKVDTT NLVNAGLTNI NLTVQ 325
<212> Type : PRT
40 <211> Length : 325
SequenceName : SEQ ID 356
SequenceDescription :

Sequence

45 <213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MKSRRQRKRG LVRKNEIIL TLFVASAVSL LAFTNSFGVL AKSLHLEKIN KSITISLPFG 60
KKKMEQTARY YSGEQVQISS SAKKDSLKGK LSHYQNWIGT VKKIKSQKDS RQKHHSYEV 120
50 TFDNGKALKY VQEKDLVTK RSKYSGQIV KLKSSATADL DGSSLTDYRA SAGKIDHISY 180
NHSNTTGGYK YDITFDEGK VTNIEKDLK KVEVQLKSE NTAAQNNEIL KQAFAYAKQH 240
SGTILSLPNG EFKIGSQTPD KDYITLSTDT EIRGDNTTLL VEGSAYWFAF ATGTSASDGV 300
KNFTMRNINI KASDLEKGNQ FMIMADHGDN WKICNNSFTM VHKKGSHIFD LGSLONSAFE 360
GNQFTGYAPE LTNVSKIDDN ADLHDFYSEV IQLDAAESSG VWDGGLIKAI DPNYENYNKE 420
55 KQLCENNITIA NNSFVPYIDS HGKIIAYSGT IGQHSSDVGL VKIYDNVFSN SLVSRFNQNG 480
KSEAWIFKAI HLKSNYNNAV YANSIS 506
<212> Type : PRT
<211> Length : 506
SequenceName : SEQ ID 357
60 SequenceDescription :

Sequence

65 <213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MRKLKVALFA SSILGMLAVS SYTAADTEDN QVTISHYNEQ AGTFDVAVQ AANGKTIQSI 60
DVAIWSEENG QDDLKWHAS NDGSNQLTVH FNAENHGSKV GSYIAHAYIT YTDGNRVGVN 120

LGKRKLSLSA PQLSLKQGG LQFSKLKPSA ADQLFSAVWS DENGQDDLHW YTADADGNTL 180
AGYANHKGYG TYHVHTYLKQ NGKMIPISAQ DIDIPKPKVK IQIDKINDTS YDVVVNNVPP 240
YISSVAIPVW SEQNGQDDLK WYQATKVADG IFKTTVYLKT HRFELGNYQA HIYGDSQLSK 300
KLDGLGETHF NVPSTIINYED PQVTIDHYNI NKGTDFVTVA ETDNSKAIQS ISAAVWSDAN 360
5 QANLYWYEAK QLANGKAAIT VDVQKHGNQT GSYNVHVYVH YNDGTTSGHV LANQQLNQIV 420
HYQPSAVRIT AYMNEKNTYP VGQCTWGVKE LAPWIPNWLG NGGQWASTVA VKGFKIGTVP 480
KVGAIACWSD GGYGHVAVYT HVESNNRIQV KEANYKNQQY ISNFRGWFD P TTSYLGRLTY 540
IYPD 544
<212> Type : PRT
10 <211> Length : 544
SequenceName : SEQ ID 358
SequenceDescription :

Sequence

15 <213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MANNYSRRQQ PTKKTGTSR KRPTEHIKTG FSALQKSAI IAGILGIITA LITINNYRNS 60
SHNDKKDSTS KTTIIEKEV DDSNSNNNA NSQAENDSNN NNNSAESNQ N QTATTANDSN 120
20 SNSANQNQAN SQSQANNQON QNNANAGQ 148
<212> Type : PRT
<211> Length : 148
SequenceName : SEQ ID 359
SequenceDescription :

25 Sequence

<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
30 MKIFSGFTIR NNTALKPNYD DTAFSGFGT IRNNTALKQS TNCASWFNRF GTIRNNTALK 60
LTILINGVSF CFGTIRNNTA LKPRGPIFVS TFRNRAIHLS QISASK 106
<212> Type : PRT
<211> Length : 106
SequenceName : SEQ ID 360
35 SequenceDescription :

Sequence

<213> OrganismName : Streptococcus mutans UA159
40 <400> PreSequenceString :
MKRKRNLVFL IGLFLTVFLL IGCSMQKKTK SESSSTSQKT TLQTKQSSEK STDAKQTTEA 60
HSESSQSSSH SNNEETLAPI DTGAVLKADY SSMAGTWKNE EGQTLTFDQR GLTTPGMTVS 120
LLNIDQDGNL LLNVETGTTK NLTLYIVPAN KTLNQNQYFSN GQSDSDKTK DRIVSSES LN 180
SGKFTNRVYY HVSTH 195
45 <212> Type : PRT
<211> Length : 195
SequenceName : SEQ ID 361
SequenceDescription :

50 Sequence

<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
55 MTPKKIKIAL TALISLMLAL FLFLFNHHSV RENSQQEKLK ISKASSKKSQ TSTSSVMTSS 60
RKATEQTSQA QTQSQSQA EQ SNP NVILPIP QELVGTYKGS SPQASEITFT ISSNGQLRAQ 120
ANFDPASDIN DVTATVSGVR KVGADTYIWE FVSGSSAALL PGVTGIGGLG KMQPGFILKG 180
GQLTPIMFTG SVDGEIDYSH PNPYPVSLNK Q 211
<212> Type : PRT
<211> Length : 211
60 SequenceName : SEQ ID 362
SequenceDescription :

Sequence

65 <213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MKKIINVIVL SLSVFFLIAC SNSSTGEKTS QSSEETKVRL IVKTDNKT D EKVAFKKGAT 60

VMDVLKDNKYK VKESGGFITT IDGVTQDKKA GRYWMFDVND KLASKAADKI KVKNGDKIEF 120
YLKVKYGKN 129
<212> Type : PRT
<211> Length : 129
5 SequenceName : SEQ ID 363
 SequenceDescription :

Sequence

10 <213> OrganismName : Streptococcus mutans UA159
 <400> PreSequenceString :
MSNKPWEEKV TDATTDNEEM TRNSKDASII STPILTILLS LFFLIIGIL FFVLYTSNGG 60
SNEKAATSGF YSSSKTVKKA KNEANSQTDE QTTEAETSSS ETTSSSSSDSD GETITVQGGE 120
GAAAIARAG ISVDKLYELN PEHMTGHWY ANPGDNIKIK 160
15 <212> Type : PRT
 <211> Length : 160
 SequenceName : SEQ ID 364
 SequenceDescription :

Sequence

20 <213> OrganismName : Streptococcus mutans UA159
 <400> PreSequenceString :
MPDNRMYISI DSNMQFPLVE ITLETGEFAY IQRGSMVYHT PSVTLNNTKVN GRGSGLGKLV 60
25 GAIGRSVTSG ESFFITQAVS NASDGKLALA PSMPGQVIAL ELGEKQYRLN DGAFLALDGS 120
AQYQMAQSV GRALFGGQGG LFVMTTEGQG TLLANSFSGI KKIQLNQEI TIDNAHVVAW 180
SRDLNVDIHL ENGFMQSIGT GEGVVNTFRG TGEIYVQSLN LQQFAGVLQG FITNTNR 237

 <212> Type : PRT
30 <211> Length : 237
 SequenceName : SEQ ID 365
 SequenceDescription :

Sequence

35 <213> OrganismName : Streptococcus mutans UA159
 <400> PreSequenceString :
MKKNYPWYGL LGLLALYLIT IAFIPGFHIF FSNMLMLALF FMLIALSNRS IFFFFLALGF 60
LSIYLDIFH FDYSTGPLFT GIIIGVILN SELKPHYSYS YKGNHYFNMK QHANYIDNET 120
40 DVFLKTLFSE NTSYVTSQEL NKIIIDTKFG EQSVLDSQAQ FMTDSPEIHI DVSFGETNLR 180
IPNNWKLINK THSPFASISF SGFPSTNGDF INVTLTGTV MGSLLNIQY 228
 <212> Type : PRT
 <211> Length : 228
 SequenceName : SEQ ID 366
45 SequenceDescription :

Sequence

50 <213> OrganismName : Streptococcus pneumoniae R6
 <400> PreSequenceString :
MKSITKKIKA TLAGVAALFA VFAPSFVSAQ ESSTYTVKEG DTLSEIAETH NTTVEKLAEN 60
NHIDNIHLIY VDQELVIDGP VAPVATPAPA TYAAPAAQDE TVSAPVAETP VVSETVVSTV 120
SGSEAEAKELE IAQKESGGSY TATNGRYIGR YGSWTAANKF WLNNGWY 167
 <212> Type : PRT
55 <211> Length : 167
 SequenceName : SEQ ID 367
 SequenceDescription :

Sequence

60 <213> OrganismName : Streptococcus pneumoniae R6
 <400> PreSequenceString :
MKHSHKKSFD WYSMQQRYSI RKYYFGAASV LLGTALVLGA AASVQTVQAE ENKQETNSI 60
SVGRGEAATK PAEVSASNKE KTYAAPTVAN PVETTPVKTE EVTKPAEKVE EAKDKKEEV 120
65 HQDAVDKSKL LTALSRAKKL ESKLYTEASA ANLQTSIQAG QSLLGKADAT EAELSAAESS 180
IQSFIIGLEL RSNNSKETVS ETPVAKKADA VESKEGAKPA ATTERSADVS AILPTSTADK 240
VETTSAPASI NEILKLGLSL SDARQNPAIR KEDVNRGYSG FRAASNPNP IVSGSGNTVA 300

	FADISQGGRS	YSFRGYGNSR	GGNSIHVDVT	TVRSGNSVNF	TISYSAPGDS	REFVNNNFIL	360
	DKGDGFGNPS	NATITSSNPR	VREQSKSISQ	GANYVSHSGY	SMTSAISTNT	EQTIRFSLPI	420
	INLNGDLSVR	LKPVTFFNDQ	GGGGAATSND	PYSNSNYYYR	ANPLYLDANP	YGGTNNKTVS	480
	EDIDFTVYL	PTSKLPEGQT	RLVREGEKGQ	RQITYKVHRF	GNETLLGLPI	SNSVTKEAKP	540
5	RIMQIGVAKD	LIDTVKPRVD	QNKVGDITNNL	TFYLDNDGNG	VYTEGVDELV	QKIAIKDGAK	600
	GEKGDQGERG	LITGAKGEKGD	RGERGLTGAQ	GAKGEKGDG	ERGLTGAQGA	KGEKGRGER	660
	GLTGAQGAKE	EKGDRGERGL	TGAQGAKEK	GAQGERGLTG	AQGAKEKGD	QGERGLTGAQ	720
	GAKGEKGDQG	ERGLTGAQGA	KGEKGAQGER	GLTGTQGAKE	EKGDRGERGL	TGAQGAKEK	780
	GDRGERGLTG	AQGAKEKGA	QGERGLTGAQ	GAKGEKGDQ	ERGLTGAQGE	KGDRGERGLT	840
10	GAKGEKGDQG	ERGITGAKGE	KGAQGERGLT	GAKGEKGEK	DQGERGLTGA	QGEKGAQGA	900
	GRDGVTPTVT	VKDNKNDGTH	TITINDGRGN	VTSTVVRDGF	DGASPLVATQ	RNDADKTTTV	960
	IFYYDKNGNN	ELDASDKKLK	EVVIADGAKG	EKGDKEQGL	QGRDGEQGP	GEDGKTPTVK	1020
	VTDGQDTH	ITINDGKGI	TTTVVRDGF	GASPLVSTHR	NEADKTTTVI	FYYDLNDNNQ	1080
	FDEGDTLKE	VVIADGKQGP	KGDKGDMGFD	GFTPEVTVTD	NNNGTHTITI	TQPDNRPSLT	1140
15	TIVKNGBDGK	TPKVKAERDD	AKKQTTLTFF	IDKDGDSYT	AGKDELQTT	VVKDQDGA	1200
	GASGRDKEV	LNGKVDPTTE	GKDGDTFVNT	QTGDVFKKG	NTWEPAGNIK	GPKGDKGADG	1260
	AKGEKGAQGE	KGLTGAQGVK	GEKGDQGERG	LTGSKGEKGD	QGERGLTGAQ	GAKGDKGEQG	1320
	LQGRDGAQGP	KGADGQRGPA	GPQGPKEQG	NPQTPGKDGK	SLIAVKNVGL	VTITPVEGRP	1380
	QTTTFVEDGQK	GADGKTPTVT	ITEGQNGTHT	LTVHNPSPD	VTTTIRDGAT	GQAGRDGKDV	1440
20	LNGKVNPPQN	QKNGDKYIN	IETGDVYVKN	NGNWDKEGNI	KGPKGDKGAD	GAKGEKGDQ	1500
	ERGLTGAQGA	KGADGAVGRD	GRDGKDVNLG	KANPEAHQK	DGDYVNTET	GDVFKNNGN	1560
	WDKEGNKGP	KGDKGADGAK	GEKGRGERG	LTGAQGAAGA	DGAAGRDRD	GRDGKDVNLG	1620
	KVNPEANQK	DGDYVNTET	GDVFKNNGN	WDKEGNKGS	KGDKGERGED	GKTPEVTVTP	1680
	GKDGHSITIT	FTVPGKDPVT	VNVKDGENGL	NGKTPKVDLL	RVQKNGNPS	HTIVTFYTDE	1740
25	NNDGKVTFTG	DELGLSEMIK	DGAKGADGRD	GKSLTLVKDG	KETKVYQEDP	ANPGQPLNPE	1800
	KPLAVIRGV	DGKSPTVTAV	RKDEAGHKGV	EITVDNHDGS	QPTTVFVQDG	AKGKTGATGQ	1860
	DGQTPITITQ	RQDQGSTVV	TITTSKDPV	TFTVKDGKNG	KDGRAPKIKV	EDITSPSRIR	1920
	RDTDAATPT	RNGIRVTVYD	DVNDNGVYDE	GVDKVLNSKD	IYNGIDGRDG	SAPTITTKDN	1980
	GDGTHITVQ	NPDGSESTTV	VKDGDGKTA	NITTTENPDG	SHTITVTNPD	GSTKETVVK	2040
30	GKDGKTPKVE	VTDNNDGTH	VKVTDDGDNV	TNAIKDGKD	GKAATATTTE	NPDGSHVTI	2100
	TNPDGKTNEF	VVKNGRDGVD	GRTPTASVRD	NGDGSHTIVI	TNPEGVTET	TVRDGKSPKV	2160
	TITDEQNGTH	KISVLNGDGT	TTETIIKDGK	SPVATVRDNQ	DGTYTIRVEN	GNGTVSETTV	2220
	RDGKSPAKV	VNDGDGTHI	TVVNSDGITT	TTTVRDGREG	KLEVIDNNDG	SHTIKVTGAD	2280
	GKGTITITFD	GKSPKANIVD	NGDGTHLTI	VDSGREGYKS	IKDGKDGKD	SVSPTVTVK	2340
35	NNDGTHVVTI	TNPDGSKTEM	VIKDGKDGKS	PKVSVEDNGD	GSHTITIINS	DGTVTKTVIK	2400
	DGKDGDRGRD	GRDGKDGKDG	KCGCQDKPVT	PSNDKVPVPT	PNVPTPEVPV	KPVPAQPTPN	2460
	VPTPEVPVQ	TPAVSTPEVP	VKPVPAVPEQ	PVPTPAQPA	TPVNPANPVAP	TTGKENRGDK	2520
	LPETGSQSDY	ISVLLGSGIL	LSLYVGRRKE	D			2551
	<212> Type : PRT						
40	<211> Length : 2551						
	SequenceName : SEQ ID 368						
	SequenceDescription :						
	Sequence						
45	-----						
	<213> OrganismName : Streptococcus pneumoniae R6						
	<400> PreSequenceString :						
	MKKRMLLAST	VALSFAPVLA	TQAEVLWTA	RSVEQIQNDL	TKTDNKTSYT	VQYGDITLSTI	60
	AEALGVDVTV	LANLNKITNM	DLIFPETVLT	TTVNEAEEVT	EVBIQTPQAD	SSEEVTTATA	120
50	DLTTNQVTV	DQTVQVADLS	QPIAEAPKEV	ASSSEVTCTV	IASEEVAPST	GTSVPPEEQA	180
	ETSSAVAEAA	PQETTPAEKQ	ETQTSPPQAS	AVEATTSSE	AKEVASSNGA	TAAVSTYQPE	240
	ETKIISTTYE	APAAPDYAGL	AVAKSENAGL	QPQTAAFKEE	IANLFGITSF	SGYRPGDSGD	300
	HGKGLAIDFM	VPERSELGDK	IAEYAIQNMA	SRGISYIIWK	QRFYAPFDSK	YGPANTWNPM	360
	PDRGSVTENH	YDHVHVSMMG					380
55	<212> Type : PRT						
	<211> Length : 380						
	SequenceName : SEQ ID 369						
	SequenceDescription :						
	Sequence						
60	-----						
	<213> OrganismName : Streptococcus pneumoniae R6						
	<400> PreSequenceString :						
	MTILGKDTVQ	QSAKGESVTQ	EATPEYKLEN	TPGGDKGNT	GSSDANANEG	GGSQAGGSAH	60
65	TGSQNSAQSQ	ASKQLATEKE	SAKNAIEKAA	KNKQDEIKGA	PLSDKEKAEL	LARVEAEKQA	120
	ALKEIENAKT	MDVKEAETI	GVQAIAMTV	PKRPVAPNAA	PKTTSAPQAT	AGTMQDVTYQ	180
	SPAGKQLPNT	GSASSAALAS	LGLVVATSGF	ALLGRKTRRR	K		221

<212> Type : PRT
<211> Length : 221
SequenceName : SEQ ID 370
SequenceDescription :

5 Sequence

<213> OrganismName : Streptococcus pneumoniae R6
<400> PreSequenceString :

10 MMTTGCSMGA YHALNFFLQH PDVFTKVIAL SGVYDARFFV GDYYNDDAIY QNSPVDYIWN 60
QNDGWFIDRY RQAEIVLCTG LGAWEQDGLP SPYKLKEAFD KKQIPAWFAE WGHDDVAHDWE 120
WWRKQMPYFL GNLVL 135

<212> Type : PRT
<211> Length : 135

15 SequenceName : SEQ ID 371
SequenceDescription :

Sequence

<213> OrganismName : Streptococcus pneumoniae R6
<400> PreSequenceString :

20 MNKGLFEKRC KYSIRKFSLG VASVMIGATF FGTSPVLADS VQSGSTANLP ADLATALATA 60
KENDGHDFFA PKVGEDQGSF EVDGPKTEE ELLALEKEKP AEEKPKEDKP AAAPETPKT 120
VTPWQTVVEK KEQQTGTVIR BEKGVRYNQL SSTAQNDNAG KPALFEKKGL TVDANGNATV 180
25 DLTFKDDSEK GKSFRGVFLK FKDTKNNVVFV GYDKDGWFWE YKSPTTSTWY RGSRVAAPEP 240
GSTNRLSITL KSDGQLNASN NDVNLFDTVT LPAAVNDHLK NEKKILLKAG SYDDERTVVS 300
VKTDNQEGVK TEDTPAEKET GPEVDDSKVT YDTIQSKVLK AVIDQAFPRV KEYSLNHGT 360
PGQVQQFNQV FINNHRTPE VTYKKINETT AEYLMKLRRD AHLINAEMTV RLQVVDNQLH 420
FDVTKIVNHN QVTPGQKIDD ERKLLSSISF LGNALVSVSS DQTGAKFDGA TMSNNTHVSG 480
30 DDHIDVTNPM KDLAKGYMYG FVSTDKLAAG VWSNSQNSYG GGSNDWTRLT AYKETVGNAN 540
YVGHSSEWQ WEKAYKGIVF PEYTKELPSA KVVITEDANA DKKVDWQDGA IAYRSIMNPN 600
QGWWKVKDIT AYRIAMNFGS QAQNPFLMTL DGIKKINLHT DGLGQGVLLK GYGSEGHDSG 660
HLNYADIGKR IGGVEDFKTL IEKAKKYGAH LGIHVNASET YPESKYFNEK ILRKNPDGSY 720
SYGWNWLDQG INIDAAYDLA HGRLARWEDL KKKLGDGLDF IYVDVWGNQ SGDNGAWATH 780
35 VLAKEINKQG WRFAIEWGHG GEYDSTFHHW AADLTGYGYT NKGINSATR FIRNHQKDAW 840
VGDIRSYGGA ANYPLGGYS MKDFEGWQGR SDYNGYVTNL FAHDMVTKYF QHFTVSKWEN 900
GTPYTMNDNG STYKWTPEMR VELVDADNNK VVTRKSNDV NSPQYRERTV TLNLRVIQDG 960
SAYLTPWNWD ANGKLLSTDK EKMYFNTQA GATTWTLPSP WAKSKVLYLK LTDQKTEEQ 1020
ELTVKDGKIT LDLLANQPYV LYRSKQTNPE MSWSEGMHIY DQGFNSGTLK HWTISGDASK 1080
40 AEIVKSQGAN DMLRIQGNKE KVSILTQKLTG LKPNTKYAVY VGVDNRSNAK ASITVNTGEK 1140
EVTYTNKSL ALNVVKAYAH NTRRNATVD DTSYFQNMVA FFTGSDVSN VTLTSLREAG 1200
DEATYFDEIR TFENNSSMYG DKHDTGKGTG KQDFENVAQG IFPFVVGVE GVEDNRTHLS 1260
EKHDPYTQRG WNGKKVDDVI EGNWSLKTNG LVSRRLVYQ TIPQNFREFA GKTIRVTFEY 1320
EAGSDNTYAF VVGKGEFQSG RRGTAQSNLE MHELNPNTWD SKKAKKATFL VTGAETGDTW 1380
45 VGIYSTGNAS NTRGDSGGNA NFRGYNDMM DNLQIEETL TGKMLTENAL KNYLPTVAMT 1440
NYTKESMDAL KEAVFNLSQA DDDISVEEAR AEIAKIEALK NALVQKKTAL VADDFASLTA 1500
PAQAQEGLAN AFDGNLSSLW HTSWGQDVG KPATMVLKEA TEITGLRYVP RGSNGNLR 1560
DVKLVVTTDES GKEHTFTATD WPDNNKPKDI DFGKTIKAKK IVLTGKTYG DGGDKYQSA 1620
ELIFTRPQVA ETPLDLSGYE AALAKAQKLT DKDNQEEVAS VQASMKYATD NHLLTERMVE 1680
50 YFADYLNQLK DSATKPDAPT VEKPEFKLSS VASDQKTPD YKQETARPET PEQILPATGE 1740
SQFDALFLA SVSLALSALF VVKTKKD 1767

<212> Type : PRT
<211> Length : 1767
SequenceName : SEQ ID 372
SequenceDescription :

55 Sequence

<213> OrganismName : Streptococcus pneumoniae R6
<400> PreSequenceString :

60 MKLYNKSELR YSRIFDKRP PAFAFILIIS TAILSGALV GAAYIPKNIY VKANGNSVIT 60
GTEFLSAISS GKVVTLHKSE GDMVNAGDVI ISLSSQEGE QASSLNKQLV KLRKEAIFQ 120
KFEQSLNEKY NRMSNSGEEQ EYVGKVEYYL SQLNSENYNN GTQYSKIQDE YTKLNKITAE 180
RNQLDADLQT LQNELIQLQQ QGDSPLSDT TSADKAKLE TKILEITTKI EALKTNITSK 240
65 NSEIDSQSN IKDMNRTYND PTSQAYNIYA QLVSELTAR SNNKKSITEL EANLGVATGQ 300
DKAHSILAPN EGTLLHVLPL KQMSIQQQG TIAEVSGKEK GYYVEAFVLA SDISRVSKGA 360
KVDVAITGVN SQKYGTLLKGQ VRQIDSGTIS QETKEGNISL YKVMIELETL TLKHGSETTV 420

LQKMPVEVR IVYDKETYLD WILEMLSPKQ 450
<212> Type : PRT
<211> Length : 450
SequenceName : SEQ ID 373
SequenceDescription :

5 Sequence

<213> OrganismName : Neisseria meningitidis Z2491
10 <400> PreSequenceString :
MNKGLHRIIF SKKHSTMVAV AETANSQGGK KQAGSSSVSVS LKTSGLDLCGK LKTTLKTLVC 60
SLVSLSMVLP AHAQITTDKS APKNQOVVIL KTNTGAPLVN IQTPNGRGLS HNRYTQFDVD 120
NKGAVLNDDR NNNPFLVKGS AQLILNEVRG TASKLNGIVT VGGQKADVII ANPNGITVNG 180
GGFKNVGRGI LTIGAPQIGK DGALTGFQDVR QGTLTVGAAG WNDKGGADYT GVLARAVALQ 240
15 GKLQKKNLAV STGPQKVDYA SGEISAGTAA GTKPTIALDT AALGGMVADS ITLIANEKGV 300
GVKNAGTLEA AKQLIVTSSG RIENSRIAT TADGTEASPT YLSIETTEKG AAGTFISNGG 360
RIESKGLLVI ETGEDISLRN GAVVQNNNGSR PATTVLNAGH NLVIESKTNV NNAKGSANLS 420
AGGRTTINDA TIQAGSSVYS STKGDTTELGE NTRIIAENVV VLSNGSIGSA AVIEAKDTAH 480
IESGKPLSLE TSTVASNIRL NNGNIKGGKQ LALLADDNIT AKTTNLNTPG NLVHTGKDL 540
20 NLNVKDLASA ASIHLSKSDNA AHITGTSKTL TASKDMGVEA GLLNVTNTNL RTNSGNLHIQ 600
AAKGLQLRN AELKPLAGRL TTALQGNIVS DGLHAVSADG HVSLLANGNA DFTGHNTLTA 660
KADVNAGSVG KGRLLKADNTN ITSSSGDITL VAGNGIQLGD GKQRNSINGK HISIKNNGGN 720
ADLKNLNVHA KSGALNIHSD RALSIENTKL ESTHNTLHNA QHERVTLNQV DAYAHRHLSI 780
TGSQIWQNDK LPSANKLVAN GVLALNARYS QIADNTTLRA GAINLTAGTA LVKRGNNINWS 840
25 TVSTKTLEDN AELKPLAGRL NIEAGSGTTL IEPANRISAH TDLSTKTGGK LLLSAKGGNA 900
GAPSAQVSSL EAKGNIRLVT GETDLRSGSKI TAGKNLVVAT TKGKLNIEAV NNSFSNYFPT 960
QKAAELNQKS KELEQQIAQL KKSPPKSKLI PTLQEERDRL AFYIQAINKE VKGKPKPGKE 1020
YLQAKLSAQN IDLISAQGIE ISGSDITASK KLNHLAAGVL PKAADSEAAA ILIDGITDQY 1080
EIGKPTYKSH YDKAALNKPS RLTRGTGVSI HAAAALDDAR IIIIGASEIKA PSGSIDIKAH 1140
30 SDIVLEAGQN DAYTFLKTKG KSGKIIRKTK FTSTRDHLIM PAPVELTANG ITLQAGGNIE 1200
ANTTFNAPA GKVTLVAGEE LQLLABEGIH KHELDVQKSR RFIGIKVGKS NYSKNELNET 1260
KLPVRVVAQT AATRSWWDTV LEGTEFKTTL AGADIQAGVG EKARVDAKII LKGIVNRIQS 1320
EEKLFTNSTV WQKQAGRGST IETLKLPSFE SPTPPKLSAP GGYIVDIPKG NLKTEIEKLS 1380
KQPEYAYLKQ LQVAKNINWN QVQLAYDRWD YKQEGLTEAG AAIIALAVTV VTSGAGTGAV 1440
35 LGLNGAAAAA TDAAFASLAS QASVSFINNK GDVGKTLKEL GRSSTVKNLV VAAATAGVAD 1500
KIGASALNNV SDKQWINNLT VNLNAGSAA LINTAINGGS LKDNLDGAAL GAIVSTVHGE 1560
VASKIKFNLS EDYIITHKIAH AIAGCAAAAA NKGKCQDGA I GAAVGEIVGE ALTNGKNPAT 1620
LTAKEREQIL AYSKLVAGTV SGVVGGDVNT AANAAKVAIE NNLLSQEEYA LREKLIKKAK 1680
GKGLSLDWG SLTEQEARQF IYLIEKDRYS NQLLDTRYQKN PSSLNQEKI ILAYFINQTS 1740
40 GGNTAWAASI LKTPQSMGNL TIPSKDINNT LSKAYQTLRS YDSFDYKSAV AAQPALYLLN 1800
GPLGFSVKAA TVAAGGYNIG QGAKAISNGE YLHGTQVQVN GTLMVAGSVS AQAATSAKPA 1860
PVTRYLSNDS APALRQALTA ESQRIRMKLP BEYRQIGNLA IAKIDVKGLP QRMEAFSSFQ 1920
KGEHGFISLP ETKIFKPISV DKYHNIAAPP RGTLRNIDGE YKLETTIAQQ LGNNNRNVSGR 1980
IDLFTELKAC QSCSNVILEF RNRYPNIQLN IFTGK 2015
45 <212> Type : PRT
<211> Length : 2015
SequenceName : SEQ ID 374
SequenceDescription :

50 Sequence

<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MDLIQTPNKQ FVDGDRRTPG TPVPAWNLNQ LQGEYSILN AVGIEPNKAD HAQVLSAIKT 60
55 LAADASQVAS IDALRKYSST GYVNVNAYHA NTTVGGGVFV ADKADKSTAD NGCTVIVSTD 120
GTRWRKVFSG MLNLHDFGYV ASKNNALSTL NAAESAALDV VVDCLGLSID TGNIYPQKNK 180
YTNGKFVING KTVDVQYQPI RSGIGRFISG TGAAANLKS EWTGAGLVI GEGAMEQMEK 240
CVSSIAIGDR AQGFSKVS RD NIAIGADSLI NVQAATEWYD QSRMEGTRNI GIGGNAGRGI 300
TSGYSNVSIG RNAGQGLGEG SSNIALGAGA MAGTAPVGFS GDIEVFWPSS TSRTIAIGEA 360
60 VLQTYQGRAA QTAIGANAAR NTKKAEKVTA IGSAAMENLE RNAPNGGDV VWTGTEAGTY 420
AQSGKNITLT FPNIRGAQAT YWVGIRLTSG TAQTLQNDVV PAQVSVNGN TLIIQSSKEL 480
TATGAELKY VYSVNSTATK NEELTIIGAN AMNKALTAGY STIIGVDAAL LGDNYQKTTA 540
IGASSLRTGS HISTTAIGYW VIPLASSEKC VAIGDSAGYR NVQGDFTLTK ITNSIAIGYG 600
ARINDNEIQ IGTGQTLYA PTAVNIRSDG RDKADV KPLT NGLDFVMKLG PMTGYD RRD 660
65 SYVDELFPKDL PADERADKVR EWWANPIKDG SHKEDRLRHV FIAQDIAALE DEYGRPLPMVN 720
KTNDTYTVEY ETFIPVLTKA IQEMAARIET LETEMKESKK 760
<212> Type : PRT

<211> Length : 760
SequenceName : SEQ ID 375
SequenceDescription :

5 Sequence

<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
MKNISRKCFM TSVVCIILGG ILLGAGYATG GLQDIKHQTA PKKVIKTFDQ ITALDIDSSA 60
STITVETGPV QRPTVTTYTH PKFIDPIVTT LTGKTLKLSQ KPKDIVITGG IEILGFTLNN 120
SRQEKNYRSI TITVPEKTSI NEVKGSNVPH TTLSNLTVDQ MQFDGNLTLL HTKVKKATIT 180
GMLEATKSQL TNLELKADYS FSNLTDSSVE NGTISLNGQ LTTKDTTLKA INIQSLHPGG 240
IEAERTTLEN VTFTVSKSKE EEEENDYYDN DAIFTAHALT LKGTNTISGG DIDVDITLTK 300
AKAIAYRART ENGKVSLSQ LTPAKIGKES TSDVTSYVAE NKAATGNLTV NLNKGDTIK 360

<212> Type : PRT
<211> Length : 360
SequenceName : SEQ ID 376
SequenceDescription :

20 Sequence

<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
MFKKENLKQR YFNFGALVALA LTILAIIFAF SSKNADTKSY AKKSESKMVT IDKAPKNNHA 60
ITKEESKEKA KSIASEPIPT VENSVAPTVT EEAPVVQEV TQTVQQVSSV AYNPNNVLS 120
NGNTAGIVGS QAAQMAAAT GVPQSTWEHI IARESNGNPN AANASGASGL FQTMPGWGST 180
ATVEDQVNA LKAYSAGLS AWGY 204

<212> Type : PRT
<211> Length : 204
SequenceName : SEQ ID 377
SequenceDescription :

35 Sequence

<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
MLEELKTLIK NPKLMITMIG VALVPALYNL SFLGSMWDPY GRVNDLPPIAV VNHDKPAKRA 60
DKSLTIGNDM VDKMSKSKDL DYHFVSSKSA QKGLKKGDY MVITLPEDLS QRATTLNPE 120
PQKLTRYQT SKGHGMVAAK MGETAMAKLK ESVSQNIKT YTSVAVSSMT DLQSGKLEAS 180
TGSQALDSGA KTAQMGSQL SDNLAGLSSA SWQFQQGTNR LTSGLTAYTA GVSQVKDGLG 240
QLSTDMPVYL NGVSRLSQGA SQLNQGLSQL TQSTTSSDDK AKRIQSLEVG LPVLNQGIQQ 300
LNENLSTMVQ PKLNTDELGN NLAAIAQAAQ QLLVKEAAAH KEQLAVLQAT SAYQSLTAEQ 360
QGELTAALTQ TDKGEAVAPA QTILRSVQTL STSLQSLSQE DQSKQLEQLK EAVAQIANQS 420
NQALPGASSA LTELSTGLAK VNGSLNQQL PGSNQLTTGL AQLNRYNTAI GSGVIKLESG 480
ANALSSKSGE LLDGSHQLSE GATKLDGSS QLSQGGHQLT SGLTELSTGL SILNGSLAKA 540
SQQLSLSVST DKNKAVAKP LVLNEKDKDG VKTNGIGMAP YMIAVSLMVV ALSTNVIFAN 600
SLSGRPVKDK WDWAQKFVI NGFISTMGSI VLYLAIQLLG FEARYGMETL GFIMLSGWTF 660
MALVTALVGW DDRYGSFASL VMLLLQVGSS GGSYPILSG AFFQKLHPFL PMTYVVSGLR 720
QTISLSGHIG VEVKVLTGFL LAFMVLSLLI YRPKKTV 757

<212> Type : PRT
<211> Length : 757
SequenceName : SEQ ID 378
SequenceDescription :

55 Sequence

<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
MSRDPYTYIN EHDLSFADGR FYVTFKADKS SETVRLNSSC LGNTIIKKLQ VEDDNTMHDF 60
VKPKVTTQQA FGLAQVKEL DLQLKDPKSD LWGKIKFNNK AMLVEYANKE MSSAIAQSAE 120
QILLQVKSID DERYSKFEQT LNKIKQTVKS ESVESARTQL ASMFDSRISG LDGKYSRLSQ 180
TIDSLSSRLD DGVGNYSLS QKVSGIDLRV SNAANDVSRL SQTAAQLQSQ ITNANQNYSS 240
LSQTVQQLQT TVRDNQSNAT SRINQLSDLI STKVSQGDVE TTIAQSYDKI AFAIRDKLPA 300
SKMSGSEIIS AINLDRSGVK ITGKNITLDG NSYISNAVIK DAHIANMDAG KINTGYLNAN 360
RIATEAITGE KIKMDYAFFN KLTANEGYFR TLFKIDIFAT SVQSVTLAS KITGGVLAAT 420
NGASQWDLNN ANMTFNRDAT INFNSKNAL VRKDGTHAF VHFSNATPKG YRGSALYASI 480

GITSSGDGID SASSGRFAGL RSFRYATGYN HTAAVDQTEL YGDNVLIADD FSINRGFKFR 540
PDKMEKVLDN NDLYAAVVAL GRCWGHLANV GWNTAHSNFT SAVSRELNNY ITKI 594

<212> Type : PRT
5 <211> Length : 594
SequenceName : SEQ ID 379
SequenceDescription :

Sequence
10 -----
<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
MAADGKVITIL VDVDGKQVKV LNSELDKVAK HGDKGSSSLK KFAVGAGVFK LASAAVDLVS 60
QSLGKAITRE DTLEKYPRVM KAMGHSADV ARSTDKTANG IDGLPTTLDE VVGTAQRLTS 120
15 ITKDINKSTN LTLALNNAFL ASGASSEAS RGLEQYQML SAGKVDMQAW KTLQETMPYA 180
LQQTAEAFGF AGASAQKDFY EALKNGQITF DQFSNKLIEL NDGVGGFAEL AKENSKGIET 240
SFNNIKNAIA KGVANSIKAL DDLSKAATGK GIADHFDLSK VVINASFSAI NASIKASTPL 300
FKLLFSVIGA GISVVKALSP ALVGVASGLA AMRAVNETIT MIKALNRAWV MASASMSIGA 360
TTIKTVTAVQ AVSTTMTKAD MVARLSQLGV LKASTVIYGV MTGAISLSTA ATIASTAAVT 420
20 ALKAALVALT GPVGWVVGAI GALVAVGVSL WSWLTKESE TKKLKKEQEG LVESNKQLRD 480
SVREGVQERK KGLSVKEST AAHQKLADI IKLAAKENKT AGEKQNLKKN IDQLNGSIDG 540
LNLAYDKNSN SLSHNADQIK SRISAMEAES TWQTAQNNLL NIEQKRSEVS KKLAEADLR 600
KKWNEBANVS DSVRKEKIAE LTEEEAKLKN MQTQLQEEYN KTSATQQAAA DAMAAEESG 660
SARQVIAYEN MSEAQRATID NMRTKYSSELL ETTTSIFDAI EQKTALSVDQ MNTNLEKNRA 720
25 ATEQWATNLE ILAQRGVDDQ ILEQLRRMGP EGATQTQVFF DATDAELAPL QENFRAATET 780
AKNAMGSVLD SAGVEMPEKV KGMVTNVSTG LQAEQAANF AQLGQEIPNG VSQGISQGAG 840
KASDASVKMG QEVKRSFQGE LGIHSPSRVF TEYGGHITDG LSNGVTINGTS KVMQTMQSLA 900
QQMSQKQQI VNDMRKSNQ ITDAFSTMSG PMHSHGVNAM QGLANGIYAG SGAALAAQAS 960
IAARITATIQ SALDIHSPSR VMRDEVGRFI PQGIAVGIDA DRKVIDSSMQ KLKESMTINA 1020
30 TPEIASGFGG GVAGIANQTT NNSNNSFTLN VKVDESDGNS HEKYQRLFRE FSWYIQQQQG 1080
RLGDVK 1086
<212> Type : PRT
<211> Length : 1086
SequenceName : SEQ ID 380
35 SequenceDescription :

Sequence

<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
MAKEPWEEKI VDDTIGTRTR KSRNAFISTP WLTALLSVFF VIIVAILFIF FYTSNSGSNR 60
QAETNGFYGA STHKKTRKAS NAKKTSSSST TDTTPSSEE TLASSEGTGE TLTVLAGEGA 120
ASIAARAGIS VEQLQALNPE HMTQGYWYAN PGDQVTIK 158
<212> Type : PRT
45 <211> Length : 158
SequenceName : SEQ ID 381
SequenceDescription :

Sequence
50 -----
<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
MSKRGIKIT TKTKLITASV ITLVLIITGV VLWKQQQNTL TADIAKEPYS TVSVTEGSIA 60
SSTLLSGTVK ALSEEIYIFD ANKGNDAVT VKIGDQVTQG QQLVQYNTTT AQSAYDTAVR 120
55 SLNKIGRQIN HLKTYGVPV STETNKDEAT GEETTTTVQP SAQQNANYKQ QLQDLNDAYA 180
DAQAEVNKAQ IALNDTVVIS SVSGTVVEVN NDIDPSSKNS QTLVHVATEG QLQVKGLTE 240
YDLANVKVGQ SVKIKSKVYS NQEWTKISY VSNYPESNA GSTTPAGSTG AGSSTGAAYD 300
YKIDIISPLN QLKQGFVSV EVVNEAKQAL VPLTAVIKK KKHVWYTDY ATGKAKKVEV 360
TLGNADAQQQ EIHKGVAVDG IVIANPDKNI KPDKKLEGGI SIGTNTKPEK DSQSKNKKSG 420
60 VDK 423
<212> Type : PRT
<211> Length : 423
SequenceName : SEQ ID 382
SequenceDescription :

65 Sequence

```

<213> OrganismName : Treponema pallidum
<400> PreSequenceString :
5  MLRLPTARAC ITMGTMIRHT FTHRCGALLC ALALGSSTMA ATAAAKPKKG QMQKLRQRPV 60
   WAPTGGRYAS LDGAFTALAN DASFFEANPA GSANMTHGEL AFFHTTGFGS FHAETLSYVG 120
   QSGNWDYAS MRMFFPESGF DFSTTEPVC TPA.SNPIKOR GAIGIINFAR RIGGLSLGAN 180
   LKAGFRDAQG LQHTSVSSDI GLQWVGNVAK SFTSEEPNLY IGLAATNLGL TVKVSDKIEN 240
   CTSTCEKCGC CKERCCCNKG KACCKDCDCN CPCQDCNDKG TVHATDTMLR AGFAYRPFWS 300
   FLFSLGATTS MNVQTLASSD AKSLYQNLAY SIGAMFDPFS FLSLSSSFRI NHKANMRVGV 360
   GAEARIARIK LNAGYRCDVS DISSGSGCTG AKASHYLSLG GAILLGRN 408
10 <212> Type : PRT
    <211> Length : 408
        SequenceName : SEQ ID 383
        SequenceDescription :

15 Sequence
    -----
    <213> OrganismName : Treponema pallidum
    <400> PreSequenceString :
20  MSRTFRAWQC VGALCALSP LPAYSSEGVR EVPPSQSPQV VVAYEPIRPG DQLLKIGIVA 60
   GCQLYIAGGN GTNGSSSSGT NGNGNGKLLG GGGPHLGYEY FFTKNFSLGG QVSFECYRRT 120
   GSNYYFSVPI TVNPTYTFAV GRWRIPLSLG VGLNIQSYLS KAPGLIAEA SAGLYYQYTP 180
   DWSIGIVAY TQLGDIASSP DKCRAVGLAT IDFGVRYHF 219
    <212> Type : PRT
    <211> Length : 219
25  SequenceName : SEQ ID 384
    SequenceDescription :

30 Sequence
    -----
    <213> OrganismName : Escherichia coli O157:H7
    <400> PreSequenceString :
   atgataaatt taagtaagga agcaacgggtg gggaaagcat taacccttat tgctatactt 60
   atgatgttgt cttttcctgt agcttctcaa gccggcggtat tagtcataaa aaatggaacg 120
35  gtatataacg ccaatggtgt gccagtcggt gacatcaaca aacctaacgg tagcggttta 180
   tctcataata tctgggataa cctaaacggt gataaaaatg gtgtcggttt caataatagc 240
   gctaataaat ccagtacttc acttgccgga aatattcagg gaaacagtaa tctgacctcc 300
   gggtcggcga aggtgatcct gaatgaggtt acttccaaaa atccttcaac cattaatggg 360
   atgatggaag ttgcagggga taaagcggat ctgattattg ccaaccggaa tgggtattact 420
40  gtaaacgggt gccggttcaat caatacaggt aaacttacct taaccaccgg gacgcgggat 480
   atccaggagt acaagctggc cggttactcc gtgaacggcg gtaccattac gctcggtaaa 540
   ctggataacg ccagcccgac agaaattctg tcccgtaaac tggtagttaa cggcaaaagt 600
   tctgccgatg agctgaacgt tgttgctggc aataactatg ttaatgccgc aggccagggt 660
   accggtagcg tatccgccac ggggtcccgt aacgggtaca gcgtagatgt tgccaaactg 720
45  ggcggaatgt atgcgaacaa aatcagtcgt ggtgtcagca tgcattccaa aggtaacctg 780
   aacctcggcg ttattgctgt ggggtttaat ggtgtcagca tgcattccaa aggtaacctg 840
   ttaaacagta acgcccagat tcagtcgtgca agcagcatca acctgacaac aaatgggtact 900
   ctggataaca ccaccggtac ggtgacatct gtaggcacta tctcgcttaa taccaacaag 960
   aatactatcg tgaatacccg tgcgggtaac atctctacga tgggcgatat ctacgttaac 1020
50  agcggtacga ttgacaatac taacggcaag ctgcggctg caggaatgct ggcggttgat 1080
   accaataacg ccacgctgat taactctggt aaagggagtt ctgtcgggat tgaagcgggg 1140
   ctgctggcgc tgaaaaccgg aacgctcaac aacagcaatg gtcagattcg cgggtggctat 1200
   gtgggtcttg aatccgctgc gctgaataac aacaacgggt atatccagac caccggcgat 1260
   atcgccatta tcagtaacgg taatgtggat aacaacaaag gtctgatccg ttcgtccacc 1320
55  gggcatatcg ttattggcgc ggcaggtagc gtaataatg gttcaaccaa aaccgccgat 1380
   accggcagtt ctgactctct gggcattatt gcagataccg gcgtagaaat tgggtgcgaac 1440
   aacatcaata acaacggcgg acagattgct tctaattggca acgtctccct gtcaagttac 1500
   agcacgatcg acgactatgc gggcaaaatt gctgccaaac gcaaagtgat tatcaaggga 1560
   agctctctgc gtaacgatac cggggggatc agcggttaag agggattga agtcgccgtt 1620
60  ggcggcagcc tgaccaataa tattggcgtg atcagctctg aagagggtga tatctccctg 1680
   ttagccaact ccgtggataa ccacggcggc ttcatgatgg ggcagaacat cagcatggag 1740
   tcgatgtctg cgtcaataa caacacagcg ctgtcgttgg ccagcaaaaa actgaagata 1800
   aatgcgcgcg gcagtatcga aaaccgcgat ggcaataact tcggtaatgc ttatgggtctg 1860
   tacttcggca tgctctcagca aacgggtgga atggtcggca aggaaggcat cgagctttcc 1920
65  gggcagaaca tctataacaa caacagccgt ctatcgtctg aggatgggtc tctgactctg 1980
   caggcgcaga acacggttga caacagcggt gctctggtca ccagcggggc ggatgcatct 2040
   attcaggttg gcggaacgta ttacaacaac tacgctacca cctggagtgc gggcaacctg 2100

```

5	gatatcgacg	cgaccacgct	gcaaaacagc	agcagcggtg	cgatgatcga	taacaatgcg	2160
	accgggttca	tagcatctga	taaaaacctg	tcaactggaa	tggtgaatag	ccttaccac	2220
	tacggctgga	tcagcggtaa	aggcgatggt	gatgtcacgg	tgaataacgg	caacctgtat	2280
	aaccgcaata	ccattgcggc	tgaaaagggg	ctgggatattg	ccgcgttgaa	cggtattgaa	2340
	aactggaagg	atatttctgc	tggcggcgac	ctgacgatga	acaccaatcg	ccatgtgacc	2400
	aacaactcca	acagcaatat	ggtggggcag	aattattgta	ttaacgcggt	taacgatata	2460
	aacaaccggt	gcaacattgt	cagtgaacgt	gacctgaacg	tgacgaccaa	aggcaacctg	2520
	tataactatc	tctatatggt	agggtatggg	gatatcgcat	tgtcggcaaa	tagcgtggcg	2580
	aacaataacg	cgaccatcga	agcgacaggg	gatctgatta	togattcgaa	gggtaacgtg	2640
10	ggtaacaacc	gcggtaatct	gcatgcggtg	aacggcggtg	tgtctgttaa	aggcaacaat	2700
	ctgaacaacg	ataacggtga	aattcgtggg	tatggcgatg	tcacgctggc	actgacgggc	2760
	aactacgaca	gctataaggg	ttcgtcgacc	tctgaaacgg	gcgacgtgac	tctgacggcg	2820
	aacattgtag	acaacgccta	tggtttgatt	gccgggtgaga	atgtttctgt	cgatgctaaa	2880
	tcgacgattt	acaacaacac	tgcgctgata	gcggcgaaat	aaaagctggt	tattaacgct	2940
15	ggcggcaacc	tcgaaaacgg	cgacgggaat	aacttctctg	gtaataacgg	cgcgctgttt	3000
	ggaattaccg	acaacgctgg	cgccatcgta	ggtaaagaag	gtgtcacgct	ttctgctcag	3060
	aacgtctaca	acaataacag	cagcatcctc	gctgaaaatg	gtccgcttaa	tctgctgtcc	3120
	aggggaaacc	ggtataatag	ccgcgcgctt	cttagcagtg	gggctgatgc	catcatccgt	3180
	gcggcgagga	cgttctacaa	caactatgcc	accacgtaca	gcgccggtaa	tctcgacgtt	3240
20	tatgcggcgt	cggtgaacaa	cgccagcgat	ggtcgccctg	aagacaatac	cgccacgggc	3300
	gtgattgcgt	ctgacaaaaa	cctggatctg	agcgttgata	acagtgtcac	taactatggt	3360
	tggatcaacg	gtataaggaga	tgtgcatttc	aatgttctga	aaggcacgct	gtataaacgt	3420
	aatgccatcg	cgccgggaca	cgcgctgacc	attaatgccc	tgaacgggtg	tgagaacttt	3480
	aaagacattg	tggcgggtac	tgcgctgact	attgatacgc	agaagtatgt	taccaacaac	3540
25	agcaacagta	atatgttggg	acaaaccatc	gcgatcaatg	ccgtgaatga	cattaataac	3600
	cggtgaaata	ttgtgggtga	ttattctctg	gggtttaaaa	ccaccggtaa	tatttataac	3660
	tacctcaata	tgctgagtta	tgggtgcgct	ggcgtatcgg	caaataaggt	tacgaatagc	3720
	ggtaaaagac	ctgttctcgg	tggcttctac	ggtttagcgt	tagaagcaaa	cgaaactgat	3780
	aacaccggta	ctattgtcgg	catgttaa				3807
30	<212> Type : DNA						
	<211> Length : 3807						
	SequenceName : SEQ ID 385						
	SequenceDescription :						
35	Sequence						

	<213> OrganismName : Escherichia coli O157:H7						
	<400> PreSequenceString :						
40	gtgaacacaa	tacacttgcg	ctgtctcttc	aggatgaatc	ccctggctctg	gtgcctgtgg	60
	gctgatgttg	cagcaaaagc	aaggctcgctt	aaacgctact	cagtattcac	ttttcagagg	120
	atgaaattta	tgaacaggac	cagtccttat	tattgtcgtc	gctcagtact	ttccttattg	180
	atatctgcct	tcatatatgc	ccgcgccggg	atggctgcct	tcactcctga	tggtattggg	240
	gtggtaaacg	atgagactgt	agatggcagc	caacgagtag	atgaacgagg	tacaacaaat	300
	aacactcata	ttatcaacca	tggccagcag	aatgtttatg	gcgggggtatc	taatggaagt	360
45	cttattgaa	ctgggtggata	tcaaga tgta	ggaaggcata	acaattatgt	ggggcagctc	420
	aataatacca	ccattaacgg	gggcagacag	tcaattcatg	acgggggtat	ttccacaggt	480
	acgataatcg	agagtggcaa	tcaggacggt	tataaaaggg	gtatcagcaa	tggaacgaca	540
	attaaggcgc	gtgcttcacg	cgtagaggga	gggagtgcca	atggaacact	cattgatggg	600
	ggttagccaga	tagtaaaagt	tcaagggcag	gctgatggta	caacgataaa	taagtctggc	660
50	tctcaggacg	tagtacaagg	aagtctggca	acgaacacaa	ccataaatgg	tggtcgacag	720
	tatgttgaac	agagcacagt	agaaacaacc	accatcaaaa	atggcgggtg	gcaaagagta	780
	tatgagagcc	gtcgcttgga	cacgacgatt	gaaggcggaa	ctcagtcctc	gaatagtaag	840
	tcaacggcaa	aaaataactc	gatctattct	gggtggtacg	aaattattga	taaccaccgc	900
	tctcggatg	ttattggaag	ttattccggt	ggcgtgcttg	atgttagtgg	tggtacggca	960
55	acaaatgtta	cccagcacga	tgggtgcaatt	ttaaaaaacta	acactaacgg	tacgacgggtg	1020
	agcgggtacg	atagtgaagg	tgcattctcc	atccacaatc	acgtggcaga	caatgtgttg	1080
	ctggaaaacg	gtgggtcattt	agacataaac	gcatatgggt	cggcaaacaa	gacgattatt	1140
	aaagataaag	gaacaatgtc	agtttttaacc	aattgctaaag	ctgatgcgac	ccgaatagat	1200
	aatggcgggg	ttatggatgt	tgcaggaaac	gcgacaaata	ccataattaa	tggtggcaca	1260
60	cagaatatta	ataattatgg	catagccaca	ggcaccaata	tcaacagcgg	aacgcaaaat	1320
	atcaaaagcg	gcgggaaagc	tgacacaaca	attatatcct	ccgggagccg	gcaggttgtt	1380
	gagaaagatg	gtacggcaat	tggcagcaat	attagcccg	gaggctcgct	gattgtctat	1440
	accggcggtg	ttgcacatgg	ggttaaccag	gagacgggca	gtgctttagt	tgccaacacg	1500
	ggtgcaggga	ctgatatcga	aggatacaac	aagctctctc	acttactat	taccggaggg	1560
65	gaggctaatt	atgtttgtgt	ggaaaaatacc	ggcggaactga	cggtagtggc	taaaacctcg	1620
	gcgaaaaata	ctaccattga	tgctggcggt	aagctgattg	tccagaagga	ggctaaaaac	1680
	gatagcacca	gacttaataa	tggcggcggt	ctggagggttc	aggacgggtg	tgaggctaag	1740

```

catgttgagc aacaatccgg cggcgcattha attgcttcca cgacctccgg aacactttatc 1800
gaaggaaacca acagtttatgg tgatgctttc tacatcagga attcagaagc taaaaatgta 1860
gtgctggaaa acgctggctc attaacagtc gtcactgggt cccgggcagt tgacacgatt 1920
attaatgccaa acggcaaaat ggatgtttat ggaaaagatg ttggcactgt actcaatagt 1980
5 gctggcaccc aaacaatata tgccagtgcc acttctgata aagcaaatat caaaggtggc 2040
aagcaaacgg tatatgggtt agccactgaa gcaaatatcg aaagtgggtga acaaatgtt 2100
gatgggtgggt caacagagaa aacacacatc aatgggtggca cgcaaacctg tcagaattat 2160
ggtaaggcgga tcaataccga tatcgtctct ggcctacaac aaattatggc aaacgggaca 2220
gcggaaggtt ccattattaa tggcggttca catagatgta atgagggcgg tctggctgaa 2280
10 aactcgggtg ttaatgatgg cggcacactc gatgtcgggg agaaaggcag cgcaacgggg 2340
atacagcaga gtagccaggg cgcgttgggt gcaaccacca gggcgacgag ggtcacagga 2400
acacgcgcgg atggcgtcgc gttcagcatc gagcagggtg cggcgaacaa tatcctgctg 2460
gcaaatggcg gagtgtaaac cgtggagtca gacacctctt ctgacaaaac acaggtcaat 2520
acggcgggag gaggatcggt caaaacaaaa gccaatcgga caggcagcac gctcacggcg 2580
15 ggtgaacaaa ttgtcgaggg tgtggcgaat gagacaacaa ttaacgacgg cggaatacaa 2640
acagtttccag ctaacggaga ggcaataaaa acaacgatca atgaaggcgg tacgctgaca 2700
gtcaacgata atggcaaacg gacagatatc gtccagaaca gcggtgcccg tctccagacg 2760
agcagcgcta ccggggcaga agatctccag gttcgtggcg ggacagcaat cgtctacgca 3000
ggcaatttag cgaccaatat gttgctggaa aatggcggtg atttattggg attagcagg 2880
20 accgaagctc gcgactccac ggttggcaag gggggggcaa tgcaaaacca gggtcaggac 2940
tccgccacaa aggttaactc tgggtggcaa tatacccttg ggcggtcaaa agatgagttt 3000
caggctctgg ccggggcaga agatctccag gttcgtggcg ggacagcaat cgtctacgca 3060
ggtacgctgg cggatgcacg ggtcagtggc gcgacaggaa gcctgtcgtt aatgacgcca 3120
cgggataatg ttacgccagt taaactcgaa ggggcgatcc ggattaccga tagcgcgaca 3180
25 ttaactatcg gcaatggcgt tgatacgacg cttgccgacc tgacggctgc cagccggggc 3240
agtgtctggc ttaacagcaa taattcctgt gcaggaccaa gtatagagta gtatagagta 3300
aacagtttgc tacttaacga cggtaatgtt tatttatcag cacaacacagc agcgcctgcc 3360
acaactaacg gtatatacaa tacgctgaca accaatgaac tttccggtag cggtaatctc 3420
tacctgcata ccaacgttgc aggtctctcg ggcgatcaac tggctcgtcaa caacaacgac 3480
30 actggttaatt ttaaaatctt tgttcaggat accggcgctc gtcctcagtc tgacgacgag 3540
atgacgctgg tgaaaacagg gggaggggat gcttcgtttt cgctgggcaa tactggcggt 3600
ttcgttgatc ttggggaccta tgagtatgtc ctgaaaagcg atggcaacag caactggaac 3660
ctgaccaatg atgtcaaacc caaccggat cccaacccaa atcccaacc aaatccgaag 3720
35 ccgacacccg ttccggagaa acgcatcacg ccttctaccg cagcogtact caatatggca 3840
gcaacattac cgttgggtatt tgatgctgag ctaaacagta ttcgagagcg gttgaacata 3900
atgaaagcga gtccacacaa caataatgtc tggggggcga cgtataacac ccgtaataat 3960
gtcaccaccc atgcgggggc cgggtttgag cagacgctga ccggaatgac agtggggatc 4020
gacagcccta atgatattcc tgaggggatt gcgacgctgg gcgcttttat gggttattcc 4080
40 cattcacata tcggttttga tcgcggagga catggcagtg tgggcagtta ttctctgggc 4140
ggctatgccg gttgggaaca tgaaagtggg ttctatctgg acggtgtcgt gaagctgaac 4200
cgttttgaaa gtaaacgtagc cggtaaaatg accggggtg gagccgcaa tggcagttac 4260
cacagcaacg ggctggggcg tcacattgaa accgggatgc gatttaccga tggtaactgg 4320
aacctgacgc cgtatgcacg gttaacgggg ttcaccgctg ataaccacca atatcattta 4380
45 tccaatggca tggaaatcgaa atcagtcgac acccgagta tatatcgtga actgggcgca 4440
acgctgagtt acaacatgag tctggggaac ggtatggaaa ttgagccgtg gctgaaggcg 4500
gctgtgcgca aagaatttgt cgatgataac cgggtgaagg tgaataatga cggtaatctc 4560
gtcaatgatt tgcggggcag acgtggaata taccaggcag gtattaaagc ctcatcagc 4620
agtagcttaa gcgggcatct tgggggtggg tatagccatg gtgcccgtgt ggaatccccg 4680
50 tggaaacggg tagctgggtg gaactggctg ttctga 4716

```

<212> Type : DNA

<211> Length : 4716

SequenceName : SEQ ID 386

SequenceDescription :

55

Sequence

<213> OrganismName : Escherichia coli O157:H7

<400> PreSequenceString :

```

60 atggcagtaa agatttcagg tgtactgaaa gacggcacag gaaaaccggg agagaactgc 60
accattcaac tgaaagccag acgtaacagc gccacgggtg tgggtgaacac ggtggcctct 120
gaaaatccgg atgaagccgg tcgttacagc atggacgttg agtacggtca gtacagcggt 180
attctgttgg tggaaagggtt cccgcgctca catgccggga ccatcacctg gtatgaagat 240
tctcaaccog gtacgctgaa tgattttctt ggtgccatga ctgaggatga tgtccgtccg 300
65 gaggcactgc gccgttttga gctgatgggt gaagaggttg ccgctaaccg gtccgcggtg 360
gcacagaaca cggcagccgc gaagaagtca gccagcgatg ccagcacatc agcccgtag 420
gcggcaaccc atgcgactga tgctgcggac tcagcacgag cagccagcac gtcagccgga 480

```

```

caggccgcgt cgtcgggtca gtcagcgtct tccagcgcag gaacggcctc aacaaaggct 540
actgaagcat caaaaaagtgc tgcgctgca gagtcctcaa aaagcgcggc ggctaccagt 600
gccggtgcgg cgaaaacgtc agaaacgaat gccggcagtg cacaacaatc agccgccact 660
tctgcatcca ccgcgaccac gaaagcgtca gaagctgcct cctcagccag ggatgcgtcg 720
5 gcttcaaaag aggcggcaaa atcatcagaa acgagcgcag cctcagcgcg cagtagtgca 780
gcttcctcgg caacggcggc aggcaattcc gcgaaggcgg ccaaaa cgtc tgagacaaac 840
gctaagtccct ctgaaacggc agcagaacag agtgctcccg cagcagcagg ctcaaaaaca 900
gcggctgcat tatctgccag tgcgctgca acaagtgcgg ggccaggcctc agccagtggc 960
accgcgcggc gaaaatcggc agaaagtgcc gcactgtctg cttcaa cagc cacaacgaag 1020
10 gctggcgaag gtgcgaaca ggccagcgca gcagcagtt ctgctt ccgc agcgaagaca 1080
tccgaaacga acgcgaaagc gtcggaaacc agcgcagaat cctcaaaaac ggctgccgca 1140
tcgtcagcca gttcggcggc gtcacgcgca tcatctgcgt ctgctt caaa agatgaggcg 1200
accagacaag cgtcagcagc gaagagcagc gccacgacgg catcca cgaa ggccagacag 1260
gcagctggta gtgcgacggc agcagctcag agcaaaagta cggcgggaatc tgcagcaacg 1320
15 cgcgctgaga cagcggcaaa acgggcagag gatattgcct ccgcgctggc gcttgaggat 1380
gcgagcacga cgaaaaaggg gatagtacag ctccagcagtg cgaactaacag cacttccgag 1440
tcactggcgg caacgccaaa agccgttaag gccgcgtatg agctggctaa cgggaaatac 1500
accgcacagg atgcgaacgg agcacagaaa gggatagttc agcttagcaa cgcgaccaac 1560
agcacatctg aaatgctggc ggcaacgcca aagtcggtaa aggcagccta tgacctgct 1620
20 aacgggaaat atactgtcga ggacgtacg acagcagaaa aaggaa ttgt ccagctcagt 1680
agtgcacca acagcgcctc tgaacgcctt gccgcgacac cgaaagcagt gaaagcagct 1740
aatgataatg cgaatggctc ggtaccttct gcccgttaagg tgaatggtaa ggcgctttca 1800
tcggatataa cactgacgcc gaaagatatt ggtacgttta actcaa caac aatgtcatc 1860
agcgggtggt ctgggttgggt caaattagca acggttaacca tggccacaggc gagttctgtt 1920
25 gtttcaatta cgttgattgg tggcgcggga tttaacgtgg ggtcac ctca acaggcaggt 1980
atatctgaac ttgttttgcg tgcaggtaat ggtaatccga agggga ttac tgggtcttta 2040
tggcagcgca catcgacagg gtttacaat tttgcctggg tcaata catc tgggtgatac 2100
tacgatattt acgttgcaat cggaaattat gcgactggtg taaata ttca atgggattat 2160
accagtaatg ccagcgtgac gattcatagc tcaccagcat attctgctaa taagccggaa 2220
30 ggggttaacgg acggtagcgt ttattcactc tatacgccat cagagcagtt ttatccgct 2280
ggcgacacaa tcccggtggc atcagatacc gttccgtctg gctatgccct gatgcagggg 2340
cagacttttg acaaatctgc ataccgaaa cttgcagccg cttatccgtc aggcgtgatc 2400
cctgatatgc gtggctggac gattaagggc aaacctgcca gtggtcgggc cgtattgtct 2460
caggacacgg atcgacacac cacagcgcca gcgcat ccag tacggatttg 2520
35 gggacgaaaa ccacatcgtc gtttgattac ggcactaaat ccacgaataa caccggggcg 2580
cacacgcaca gtgtgagcgg tacagccgca agtgccggaa accata ctca tagtgtcaca 2640
ggcgcatcag cagtcagcca gtggtcacaa atgggtcag tacata aggt agtgtctgcg 2700
gccagtgtag atacaagtgc tgcaggagcg cacactcata gtgtcagcgg cacagctgca 2760
tctgcagggt ctacgcgaca tactgtcggg attggtgctc atacgcactc tgttgcgatt 2820
40 ggctcacatg gacacaccat caccgttaac gctgcgggta acgcggaaaa cactgtcaaa 2880
aacatcgcat ttaactacat tgtgaggctt gcataa 2916

```

<212> Type : DNA

<211> Length : 2916

SequenceName : SEQ ID 387

45 SequenceDescription :

Sequence

```

<213> OrganismName : Escherichia coli O157:H7
50 <400> PreSequenceString :
atgaaacgag ttattaccct gtttgctgta ctgctgatgg gctggc cggc aaatgcctgg 60
tcattcgcct gtaaaaccgc caatggtacc gctatcccta ttggcgggtg cagcgctaata 120
gtttatgtaa accttgcgcc tgcggtgaat gtggggcaca acctggtcgt agatctttcg 180
acgcaaatct tttgccataa cgttatccg gaaaccatta cagactatgt cacactgcaa 240
55 cgaggctcgg cttatggcgg cgtgttatct aatttttccg ggaccgtaaa atatagtggc 300
agtagctatc catttccgac caccagcgaa acgcgcgggg ttgtttataa ttcgagaacg 360
gataagccgt ggccgggtggc gctttatttg acgctgtga gcagtgcggg cgggggtggc 420
attaaagctg gctcattaat tgcgctgctt attttgcgac agaccaaaaa ctataacagc 480
gatgatttcc agtttgtgtg gaatatctac gccataatg atgtggtagt gcctactggc 540
60 ggctgcgatg tttctgctcg tgatgtcacc gttactctgc cggactacc tggttcagtg 600
ccaattcctc ttaccgttta ttgtgcgaaa agccaaaacc tggggatta cctctccggc 660
acaaccgcag atgcgggcaa ctcgatttcc accaataccg cgtcgttttc accagcgag 720
ggcgtcggcg tacagttgac gcgcaacggg acgattattc cagcgaataa cacggtatcg 780
ttaggagcag taggaacttc ggcggttaag ctgggattaa cggcaaatca cgcacgtacc 840
65 ggcgggcagg tgactgcagg gaatgtgcaa tcgattattg gcgtgacttt tgtttatcaa 900
taa
<212> Type : DNA

```

<211> Length : 903
SequenceName : SEQ ID 388
SequenceDescription :

5 Sequence

<213> OrganismName : Escherichia coli O157:H7

<400> PreSequenceString :

10	atgggctacg ttacaggtgg attaccaatg aagaataacc gtgcgtgggc gcttatcagt	60
	gggtctgatat tgttcagcgg aacggcccca gctgcccata acctgcattt tacccggaat	120
	ttgcttggtg aatcctgtac tctgtgaatc aatggcaact tacttgacaga aattcatttc	180
	cccacaattg ctgccagcga tttaatgcaa cgtggtcagt cagatcgcg accggttagt	240
	tttcagttga aagattgcaa aagcaccacg gcgtttaatg tcaagggtgac cttgatggga	300
	acagaagata ccgacttacc aggtatttctg tgcattgatt cgtcatcttc tgcaacgggt	360
15	gttgggattg gcattgaaac tgccggaggg gcggtgtgac ctattaacag taccacaggt	420
	gcctcatttc cattaaatca gggaaataac agtgtcaatt ttaatgcctg gttacagacc	480
	gtaaattggac gaaatgttac atcgggtgat ttcaccgcca caatgacggg aacttttgag	540
	tattttttaa	549

<212> Type : DNA

20 <211> Length : 549

SequenceName : SEQ ID 389
SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7

<400> PreSequenceString :

25	atgaaacgac atctgaacac cagctacagg ctggtatgga atcacattac gggcaccctg	60
	gtgggtggcct ccgaactggc ccgctcacgg ggaacacgcg ccggtgtggc ggttgcgctg	120
30	tctcttgata ctgtcacatc agtcccggca ctggctgctg acaaggttgt acaggcggga	180
	gaaaccgtga acgatggaac actgacaaat catgacaacc agattgtctt cggtagcgcc	240
	aacggaatga ccatcagtag cgggctggaa ctggggccgg acagtgaaga aaacaccggg	300
	gggcaatgga tacagaatgg cgggatagcc ggaaacacca ctgtcaccac aaatggtcgt	360
	caggtcgtgc tggagggggg aacagccagt gatacggtta ttcgtgacgg cgggggacag	420
35	agcctgaacg gactggcggg gaacaccaca ctgaataaca gaggcgagca gtgggtgcat	480
	gagggcgggg ttgccaccgg tacaattatc aaccgcgacg gttaccagag cgttaaaagt	540
	ggcgggctgg caacaggaac catcatcaac accggcgacg aaggcggccc tgattctgac	600
	aactcgtata cgggtcagaa ggtccaggga acagcagaat ccaccaccat caacaaaaat	660
	ggacggcaga ttatcttatt ttccgggcta gccctgtgaca ctctcattta cgcaggtggg	720
40	gaccagtcgg tacacggaag ggccctgaat accacactga atggcgggta ccaatatgtg	780
	cacagggacg gacttgcgct gaacacggta attaacgagg ggggctggca ggttgttaag	840
	gcaggtggcg gcgctgttaa caccaccata aatcagaacg gtgaactgag ggtacatgcc	900
	ggcggggaag ccactgcagt caccagaac acggcgcggt cactggttac cagtactgct	960
	gcaactgtca tcggcacaac ccgtctgggg aatttcacgg tggaaaacgg taaggctgac	1020
45	ggtgtgtgtt tggaatccgg cggctgctct gatgtactgg agagccattc agcacagaat	1080
	accctagtgg atgacggcgg tacctgggca gtgtctgccc gcggtaaggc gacaagtgtc	1140
	accataacat ccggtggtgc cctgattgca gacagtgggt ccactgttga ggggaccaat	1200
	gccagcggta agttcagtat tgatggcaca tccggtcagg ccagcggcct gctgctggaa	1260
	aatggcggga gctttacggg taatgccggg ggacaggtcg gcaacaccac tgtcggacat	1320
50	cgtggaacac tgacgctggc tgcgggggga agtctgagtg gcagaacaca gctcagtaaa	1380
	ggcgccagta tggtagtgaa tggtagtggt gtcagtagcg gcgatatgtg taacgcaggg	1440
	gagattcgct ttgataatca gacgacaccg aatgcccgcg tgagccgtgc tgttgcaaaa	1500
	agtaactccc cggtaacgtt ccataaactg accaccaga acctcaccgg ccagggcggc	1560
	cccatcaata tgcgtgttcg ccttgatggc agcaatgcct ctgaccagct ggtgattaat	1620
55	ggtgggtcagg caaccggcaa aacctggctt gcgtttacaa atgtcggaaa cagcaacctc	1680
	gggggtggcaa ccaccggaca gggatatccg gttgtggatg cacagaatgg cgccaccaca	1740
	gaagaagggtg cgtttgccct gactcgcccc cttcaggccg gcgcctttaa ctacacctg	1800
	aaccgtgaca cggatgaaga ctggtagctg cgcagtgaat atgcttatcg tgctgaagtc	1860
	cccctgtata catccatgtt gacacaggca atggactatg accggattct ggcaggctcc	1920
60	cgcagccatc agaccggtgt aaacgggtgaa aataacagcg tccgtctcag cattcagggc	1980
	ggtcatctcg gtcacgataa caacggcggt attgcccgtg gagccacgcc ggaaagcagc	2040
	ggcagctatg gactcgtccg tctggagggg gacctgctca gaacagaggt tgccgggtatg	2100
	tctctgacga caggggtgta tggtagtgca ggccattctt ccgttgatgt taaggatgat	2160
	gacgggttccc gcgcccgcac ggtccgggat gatgccggca gtctgggccc atacctgaat	2220
65	ctgggtacaca catcctccgg cctgtggggt gacattgtgg cccagggaac ccgtcacagc	2280
	atgaaagcgt catcggacaa taacgaacttc cgcgcccggg gctggggctg gctgggctca	2340
	ctggaaacgg gtctgccctt cagtatcact gacaatctga tgctggagcc acaactgcag	2400

```

    tacacctggc agggactctc cctggatgac ggccaggata acgccgggta tgtgaagttc 2460
    gggcatggca gtgcacaaca tgtgctgccc ggtttccgtc tgggcagcca caa.cgatatg 2520
    acctttggtg aaggcacctc atcccgtgac accctgcgcg acagtgcaaa aca.cagtgtg 2580
    agtgaactgc cgggtgaactg gtgggtacag ccttctgtta tccgcacctt cagctcccgg 2640
5   ggtgacatga gcatggggac agccgcagcc ggacgtaaca tgacgtttct accgtcccgg 2700
    aatggcacgt cactggacct gcaggccgga ctggaagccc gtatccggga aaatatcacc 2760
    ctgggcgttc agcccggtta tgcccacagc gtcagcggca gcagcgtga aggcctataac 2820
    ggtcaggcta cgctgaatat gactttctga 2850
    <212> Type : DNA
10  <211> Length : 2850
    SequenceName : SEQ ID 390
    SequenceDescription :

Sequence
-----
15  <213> OrganismName : Escherichia coli O157:H7
    <400> PreSequenceString :
    atgaaaaaat ggcattatat attttgcata attctctttc atttaggggt accgtgcggg 60
    tatgcggcaa atgatggcac gtgtgcaaca agaggcggca cacatacatt aagccttaat 120
20  tttcctctga caacggtcag tgcagcaaac aatgtgcctg gaaatacatt aatagatatt 180
    gctaattgcaa catctttctga aaattatagc gttctgtgta actgtgattc aaa.acatagc 240
    aatggcgctt atcacgaaat atattatacc gcagaccctg ctcccgggat ggt.ttatagc 300
    accaccgcaa gtggtctttg tttttactat cttaacgaat atgtcgatgt gggaacaaaa 360
    atatctgtgc taaatgcccc gtatacggca gttccttttg aacatgtttc caa.ccaggca 420
25  actacaacag atcacacttg tcagggaaac aaaactacag cggttggcgt gagcctgaaa 480
    actggagcag atgcgaagat ttcatttcgt attaaacggt caataaatgg aacggtagta 540
    atacctatca ccgatattgc attgctgtat gccaacatat ccagcaccac gaccctgggt 600
    gaggcgattg caaaagttcg aatttcaggc agtttgaccg caccacagtc ttgtcagata 660
    aatgcaggac aggtgattta ttttgatttt gatactattc ctgcgtccga attttcatct 720
30  accgccgggc aagccattac ttcacgaaaa atcactaaaa cagttagtat tgagtgtacg 780
    gggatggggg atgagcgtac gcagaaagtc gatgcttctt ttacggggac gaaccgaaagc 840
    agtgacgata cgatgggtggc gacagacaat gctgatgtcg ggatcaaaa ttacaataaa 900
    tcgaatgctg aagttagcgt caacaacggc aggttaccgc cagacatggg caa.cacgacc 960
    atttttggtc gtaaaaatgg ttccgtaact ttttcggcag cacctgccag ctt.taccggt 1020
35  gcccgccctc agcccgcgct ttttaacgct accgcgacct taaccattga att.tgtaaac 1080
    taa 1083
    <212> Type : DNA
    <211> Length : 1083
    SequenceName : SEQ ID 391
    SequenceDescription :

Sequence
-----
45  <213> OrganismName : Escherichia coli O157:H7
    <400> PreSequenceString :
    atgtcacggt ataaaacagg tcataaacaa ccacgatttc gttattcagt tctggcccg 60
    tgcgtggcgt gggcaaatat ctctgttcag gttctttttc cactcgctgt caccctttacc 120
    ccagtaattg cggcacgtgc gcagcatgcg gttcagccac ggttgagcat gggaaatact 180
50  acggttaactg ctgataataa cgtggagaaa aatgtcgcgt cgtttgccgc aaa.tgccggg 240
    acatttttaa gcagtcagcc agatagcgat gcgacacgta attttattac cgggaatggcc 300
    accgctaaag ctaaccagga aatacaggag tggctcggga aatatggtac agcgcgcgctc 360
    aaactgaatg tcgataaaga tttctcgctg aaggattctt cgctggaaat gcttttatccg 420
    atttatgata cgccgacaaa tatgttgttc actcaggggg caatacatcg tacagacgat 480
    cgtactcagt caaatattgg ttttggctgg cgctattttt caggaaatga ctggatggcg 540
55  ggggtgaaca cctttatcga ccatgattta tcccgtagtc ataccgcgat tgggtgttgg 600
    gcggaatact ggcgcgatta tctgaaactg agcgcgaatg gttatattcg ggccttctggc 660
    tggaaaaaat cgccggatat tgaggattat caggaaacgc cggcgaatgg ttgggatatc 720
    cgcgcagagg gctattttacc tgccctggccg cagcttggcg caagcctgat gtatgaacag 780
    tattatggcg atgaagtcgg gctgtttggg aaagataagc gccagaaaga ccgcgatgct 840
60  atttctgcgc aggtgacctt tacgccagtg cctctctctga cactgagcgc cgggcataag 900
    cagggcaaga gcggtgagaa tgacactcgc tttggcctgg aagttaacta ccgaattggc 960
    gaacctttgg cgaaacaact cgatacggat agcatcgcg agcgtcgggt actggcaggc 1020
    agccgctatg acctggttga gcgtaataac aacatcgctt ttgagtaccg caaatctgaa 1080
    gtgatccgta ttgctctgcc tgagcgtatt gaaggtaaag gcggtcagac actttccctg 1140
65  gggcttgggt tcagcaaagc aactcacgga ctgaaaaatg tgcagtggga agcgcgctca 1200
    ttactggctg aaggtgggca aattaccggt cagggtagtc agtggcaagt aacgctccc 1260
    gcttatcgtc caggcaaaaga caattattat gcgatttcag cagttgccta cgataacaaa 1320

```



```

5  ggcaataacct caaaacgcgt gcagacagag gtgggtcatta ccggagctgg tatgagcgcc 1380
   gatcgcacgg cggttaacgct tgacgggtcag agccgtattc aaatgcttgc taacggtaaat 1440
   gagcaaaaac cgctgggtgct gtctctgcgc gacgccgagg gccagccagt cacgggcatg 1500
   aaagatcaga tcaagactga actaatcttc aaaccggctg gaaatattgt gactcgttcc 1560
   ctgaaggcca ctaaatacaca ggcaaaagcca acactgggtg agttcaccga aactgaagca 1620
   ggggtgtatc agtctgtctt tactaccgga acgcagtcag gtgaggcaac gattactgtt 1680
   agcgttgatg gcatgagcaa aaccgtcact gcagaactgc gggccacgat gatggatgtg 1740
   gcaaactcca ccctgagcgc taacgagccg tcaggtgacg tggttgctga tggtcagcaa 1800
   gcctatacgt tgacgttgac tgcggtggac tccgagggtg atccggtgac gggagaagcc 1860
   agccgcttgc gatttgttcc gcaagacact aatgggtgaa ccggttgggtgc catttcggaa 1920
   ataaaaccag gcgtttacag cgcgcgggtt tcttcgaccc gtgccggaaa cgttgttgtg 1980
   cgtgctttca gcgagcagta tcagctgggc acattacaac aaacgctgaa gtttgttgcc 2040
   gggccgcttg atgcagcaca ttcggtccatc accctgaatc ctgataaacc ggtggttggg 2100
   gggcagagta cggcaatctg gacggtaaaa gacgctatg acaacccctgt gaccagcctc 2160
   acgcccgaag cgccgtcatt agcgggtgccc gctgctgaag gttctacggc atcgggctgg 2220
   acaataaatg gtgatgggac gtggactgcg cagattactc tcggctctac ggcgggtgaa 2280
   ttagaagtta tgccgaagct aaatggacag aatgcggcag caaatgcggc aaaagtaacc 2340
   gtgggtgctg atgcgttatc ttcaaacacc ctgaaagtct ctgtcgcaga agatcacgta 2400
   aaagccggcg aaagcacaaac cgtgacgctg gtggcgaaag atgcgcatgg caacgctatc 2460
   agtggctctg cgttgtcggc aagtttgacg gggaccgcct ctgaaggggc gaccgtttcc 2520
   agttggaccg aaaaaggtta cggttcctat gttgctacgt tgactacagg tggaaagacg 2580
   ggcgagcttc cgtcctgccc tctctcaac ccgcagccag cagccaccga agccgcgcag 2640
   ttgacggtca tcgccggaga gatgtcatca gcgaactcta cgttgttgc ggacaataag 2700
   gctccgaccg tcaaaacgac gacggaactc accttcaccg tgaaggatgc gtacgggaac 2760
   ccggtcaccg gggtgaagcc agatgcacca gtgtttagcg gtgccgccag cacggggagt 2820
   gagcgtcctt cagcaggaaa ctggacagag aaaggtaatg ggggtctacgt gtcgacctta 2880
   acgctgggat ctgccgcggg tcagttgtct gtgatgccgc gagtgaacgg ccaaaatgcc 2940
   gttgtcagc cactgggtgct gaacgttgca ggtgacgcat ctaaggctga gattcgtgat 3000
   atgacagtga aggttaataa ccaactggct aatggacagt ctgctaacca gataaccctg 3060
   accgttgttg acacctatgg taaccggttg caggggcagg aagttacgct gactttaccg 3120
   cagggtgtga ccagcaagac ggggaataca gtaacaacta atgcggcagg taaagcggac 3180
   attgagctta tgtcaacggt tcgcccggaga cacaatatct ccgcttcggt gaatggtgct 3240
   cagaagacgg tcacgggtgaa attcaacgcg gatgccagca ccggtcaggc aaacctgcag 3300
   gtgacgcccg cgtctcaaaa agtggcacaac ggcaaagatg cctttacgct gacggcgaac 3360
   gttgaggata aaaatggtta ccctgttcca gggagcctgg tgacctttaa tctgcccccg 3420
   ggtgtcaagc cgcttacagg cgataatgtc tgggtgaaag ccaacgatga ggggaaagca 3480
   gagttgcagg tgggtttcagt gactgcgga acgttatgaga tcacggcatc ggcagggaat 3540
   agccagcctt cgaatacgca gactataacg tttgtagccg ataaggctac cgcaaccgtc 3600
   tccggtattg aggtgattgg caactatgca ctggcggacg gcaatgccaa acagacgtat 3660
   aaagttacgg tgactgatgc caataacaaac ctgttgaaag atagcgaagt gacgctgact 3720
   gccagcccg ccaatttagt tctgactccc aatgggacgg cgaaaactaa tgagcaaggga 3780
   caggctatct tcaccgccac gaccactgtc gcagcgaaat atacactcac ggcgaaagtg 3840
   agtcaggccg acggtcagga atgcagcaaa actgccgaat ctaaattcgt cgcggatgat 3900
   acaaatgcag tactaccgc atcatctgat gtgacttctc tgggtggcggg tgggatatcg 3960
   actgcgaagc ttggaggtgac actgatgtcg gcaataaacc ccgttggggg gaatatgtgg 4020
   gtcgacatta agacgccaga aggggtgacg gagaaggatt atcagttcct gccgtcgaaa 4080
   aatgaccatt tcgtgagcgg aaaaatcacg cgtacattta gtaccagcaa gcctggtgtc 4140
   tatacgttca catttaacgc cctgacgtat ggcgggtacg aaatgaagcc agtgacggtg 4200
   accattaccg cgggtggatgc cgatacggca aaggcgaggg aggcgatgaa ctaa 4254

```

<212> Type : DNA

<211> Length : 4254

SequenceName : SEQ ID 392

SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7

<400> PreSequenceString :

```

60  atggcgcggtg gttggggtgc ttcagaagcc tcaggcgcca tgactgattg gttaaataac 60
   tttgggtactg cgagaatctc tctgggtgtg gatgaagatt ttagcctgaa aaattcgcaa 120
   ttgcacttcc tgcattccgtg gtatgacaca cctgattatc tgctcttcag ccagcatacc 180
   cttcaccgaa cagacgatcg taccagatc aacaccggtt tgggctggcg tcatttcacc 240
   tocagctgga tgtcaggcat caaccttttt tttgaccacg acctgagccg ctatcactcc 300
   cgccgagggc ttggcgacga atactggcgt gattatctga agttgagcag caacgcttat 360
   atcggcctga ccggctggcg tagcgcacca gaattggata acgacttcga agcccgcccg 420
   gccaacggct gggattttacg cgccgaaggc tggttacctg cctggccaca actgggggga 480

```

	aaactgggtct	atgaacaata	ctatggcgat	gaagtggcgc	tgtttgacaa	gaatgatcgt	540
	caaagtaacc	cccatgctat	taccgcaggc	ctcaactata	cccccttccc	gcttctgact	600
	ctcagtgccg	aacagcgtca	ggggaagcaa	gggtgaaatg	acacacgttt	tgccgttgat	660
	ctgacctggc	aaccacagcag	ttcaatgcag	aaacagctta	atccgggacga	agtggccgga	720
5	cggcgccagtc	tggccggtag	tcgttatgac	ctgattgatc	gcaacaacaa	catcggtctg	780
	gaataaccgca	agaaaagcgt	gattcgccgtg	agtctgctgg	atccgggtgaa	agggaaagtct	840
	ggagaaaataa	aaccgctggg	ttcctcgcta	cagaccaaat	atgcccttaa	aggctataac	900
	atcgaagccg	ctgcgctgga	agctgccgga	ggtaaaagtc	gcacgtctgg	aaaagatac	960
	acggctcacgc	tggcaggtta	ccgcttcact	aacaccccg	aaaccgataa	tacatggctg	1020
10	atagacgtta	ccgcccaggga	tgtaaaaggt	aacctgtcac	ggcatgaaca	aagcatggta	1080
	gttatttcagg	ctccgacatt	aagccagaaa	gattctctgt	tatccgtcaa	tccgctaacc	1140
	gtggctgcag	ataaaaaatc	gacgaccaca	ttgaccgtta	ctgcgcacga	ttccgacgga	1200
	actccgggtgc	cggggctggc	gctgcaaac	cgcagtgaag	gcgttcagga	tatcacctg	1260
	tctgacgtta	agataacgg	tgatggtagt	tacacacaga	tactgaccgc	cggaaacgaca	1320
15	tcagggttcag	taacactgac	gccgcaaat	aacgggtgaga	gtgcgggtaaa	agaatccatc	1380
	gtcggttaata	tcgtccctgt	tgtctcatcc	cgcgaccatt	catcaataac	aattgataac	1440
	gtatcgattt	atgccgggaga	cgacatcaag	gttaggggtgg	aactgaaaga	cgatagcaat	1500
	caagacgttt	catatcaaaa	agaggaattg	gtaaaagccg	ttactgtcga	aaacagcaaa	1560
	cctggcgcca	cgattgtctg	gcacgaagag	cagccggggg	tttatgccgc	gaattatccg	1620
20	gcctataaagc	aagggactgc	actaagggca	caacttagcc	ttcacaactg	gaatgctcca	1680
	ctgcaatcgc	atatttataa	cattgaggca	aaccagaata	aggctcgcgt	tgccacatta	1740
	tcagtcagca	ataatgacgt	ttacgcgat	aaaaagacat	ttataaccct	cacgatcaac	1800
	gtcactgatg	agagtataa	ttccctgaca	aatcatcagg	tcacctttaa	gaatgaaaaa	1860
	ggaagcgcgg	agtttgtcga	accgcgcgag	caaaatacgg	atgcatatgg	tggtggccaca	1920
25	ataaacatgg	taagtccaggt	tgccgaagaa	aatacagatta	gcgccacgct	gccaaatggg	1980
	ttttcacatca	ggataaattgc	gaaattcggt	agcgattcga	gtacgccaaa	attcaaacaa	2040
	ctgggttgcg	atccagatac	cattattgct	ggcaacagcc	agggcagtag	tctgaccgcc	2100
	atcatcacag	actttcataa	caaccggtta	aaagatatga	aagtgaattt	tgtggcacct	2160
	gggtggctcgc	aactggacaa	cacgaccgcc	acaacagacc	agtcgggtat	tgtgcccgggtg	2220
30	cacctgaccca	gttcaaaaagc	tggtagctat	tcogtcgatg	cctcgcttga	ggtaggataaa	2280
	aatatttcacc	agtcgggtcac	gatcacccgtg	gtcccaaaaca	gggaacaatc	ggtaatgacc	2340
	ttgaatgccg	ggtcggggcag	tgcgatcgct	aacaatacaa	atatcggttac	cctgactgcc	2400
	agtggtgaaa	atgttttatgg	acaccggttg	ccggatgagg	atgtgaaatt	taccttgcca	2460
	gcctccatga	gcgggaacct	cacgctaagt	acggaacccg	cccgcaccga	tgcaaacggg	2520
35	gatgccgtgg	tcacattgcg	aggcacaaaa	gcgggtgagt	ttacagttac	ggcgacgctg	2580
	accagaaata	ataccggttg	ttatcagcaa	gtcactttta	ttggggatac	aaacagtgcg	2640
	cagctccagc	cgctgactgc	ctcattaaat	tccattgttg	cgggtaacag	tacggggagt	2700
	accctcagcg	gtataaccc	ggacgcttac	caaaatccgc	ttaaaagacca	gttgggtcact	2760
	ttccagagta	acgatgtcac	tctaagcgaa	acagaagtca	ccaccaatac	gctgggtcag	2820
40	gcgacgggtaa	caatgaccag	caatattgcc	ggacaacata	acgtcgtggg	gagccggaaa	2880
	gcgcaagctt	ccgataataa	aacgtttagt	ttatcagtg	taccggatga	aagttcggcg	2940
	agtaacagta	agcgaacccg	agccgaaaaa	acgataacgg	tgggcgaaaa	catcacgcta	3000
	cggataactcg	tccaggacgc	gtttaacaat	gtaatcgccg	gtcaacgcgt	cagattaagt	3060
	gcgcagccaa	caactaacat	tacgataggg	gatacggctt	acaccgataa	taacgggttat	3120
45	gcgtacgtta	accttctcag	cacccaacct	gggggtttatc	aggtgacggc	aacgctggac	3180
	aataacagta	gtagtaaggt	tgacgtgaa	gtggcaaatg	gcaaaactcg	gttaacatca	3240
	tcgaaaccag	aaactacggg	ccataatagt	gagggtatta	cgctgaccgc	aacggcgaga	3300
	aatgcgcggg	gtgaattgat	gccagggcaa	attatcacct	ttagcgtaac	gcctgaaggt	3360
	gcaacgctaa	gcaatacagg	ggaagtcctt	actgaccagt	caggtcaggc	caaagtgcag	3420
50	ctgaccagtg	acaaagtgaa	tgtctatacc	gttacggcca	taatgggcaa	agatgttccc	3480
	gttcagagcc	aggtaacggg	tgccgttaag	gcagatgcta	aaacggcaca	tgttgtagac	3540
	gtcgtggctt	ctcctgacac	catcacccgc	gacggcatcg	atagcagcac	catcacttca	3600
	cgagtagaag	atgattacgg	attcccgggt	gaaggtgtcg	atattagtca	tggcttagac	3660
	accaaaggca	gcccggtagt	taatatccca	actacgcgta	ccgatcagtc	cgggcaagtc	3720
55	acggcgacaa	taaccagtag	attggcagaa	accttaacag	tcaatgtgca	agttcctggc	3780
	acagccaacc	aatccgcac	cattacattg	gttgccggca	cggccgatga	aagtaagtca	3840
	attttgaaat	ccgatgttga	cactctgaag	gctgactacc	agcagagcgc	aaaacttacg	3900
	ctaataattgc	aagacaagta	cggtaacccg	atagtgacgt	ctgatcatct	ggaatttgtc	3960
	cagtcaggcc	ccttcgtgaa	ctttctcaag	ttgagcgata	ttgattacag	ccaaagaaat	4020
60	tatggcgagt	acaccgtgac	tgtcactggc	ggaaaagagg	gaacagcgac	actcattccc	4080
	atgctgaacg	gggttcatca	ggcaaaactta	agcatatcgc	tgaatctcat	ccaatcgata	4140
	aaagaaatgc	ccggtcatgt	cactgcaaac	aaccatacct	tctccacggc	taaatcccg	4200
	agcgaaggct	ttgcaggagc	gtattacaca	ctcaacaatg	ataactttga	agcgggtaaa	4260
	accgttgatg	atttatgttt	ttcaagttca	cagggttggg	tgtctgtcga	tgcttcgggt	4320
65	aaagtttctt	tcgcaaatat	cggcgatcaa	acgtcagtc	caataagcgc	tgttccccga	4380
	caaggaggtg	caacctacca	gaccttaatt	aagctgaaag	gctgggtggg	gaataatgga	4440
	aatcatacca	atatctgggt	agctgccaat	gcgctctgtc	atgctaaaaa	tgatggatat	4500

```

    aatcttctctg gcatcacaca tttgacgtct ggcgaaaaca aacgcacgca gggatcactg 4560
    tatgggtgaat gggggaacgt tggagcggtt tccagtaatt cgcaatttac accgggagct 4620
    tactggacaa gtgaatctga tgattacagt cggcactact atgtgcagat gctaaccggt 4680
    atgaccggaa gcgacgctga ttccagcccc caactgaccg cctgccgtaa atcactttaa 4740
5
    <212> Type : DNA
    <211> Length : 4740
        SequenceName : SEQ ID 393
        SequenceDescription :
10
    Sequence
    -----
    <213> OrganismName : Escherichia coli O157:H7
    <400> PreSequenceString :
15
    atgattactc atggttggtta taccgggacc cggcacaagc ataagctaaa aaaaacattg 60
    attatgctta gtgctgggtt aggattgttt ttttatgtta atcagaattc atttgcaaat 120
    ggtgaaaatt attttaaatt ggggttcggat tcaaaactgt taactcatga tagctatcag 180
    aatcgctttt tttatacgtt gaaaactggt gaaactgttg cggatctttc taaatcgcaa 240
    gatattaatt tatcgacgat ttggctcgtt aataagcatt tatacagttc tgaaagcgaa 300
    atgatgaagg ccgcgcctgg tcagcagatc attttgccac tcaaaaaact tccctttgaa 360
    tacagtgcac taccactttt aggttcggca cctcttggtt ctgcaggtgg tgttgcgtgt 420
    cacacgaata aactgagctaa aatgtcccg gacgtgacca aaagcaacat gaccgatgac 480
    aaggcattaa attatgcggc acaacaggcg gcgagtcctc gtagccagct tcagtgcgca 540
    tctctgaacg gcgattacgc gaaagatacc gctcttggtt tcgctggtaa ccaggcttcg 600
    tcacagttgc aggcctgggt acaacattat ggaacggcag aggttaattc gcagagtggt 660
    aataactttg acggttagtt actggacttc ttattacgtt tctatgattc cgaaaaaatg 720
    ctggcatttg gtcaggctcg agcgcgttac attgactccc gctttacggc aaatttaggt 780
    gcggttcagc gttttttcct tccctgcaac atgttgggct ataactggtt cattgatcag 840
    gatttttctg gtgataatac ccgttttagt attgggtggc aatactggcg agactatttc 900
    aaaaagtagc gtttaacgcta ttccgcctg agcggctggc atgagtcata caataagaaa 960
    gactatgatg agcgcaccagc aaatggcttc gatccggtt ttaatggcta tctaccgtca 1020
    tatccggcat taggcgcaa gctgatatac gagcagatatt atggtgataa tgttgctttg 1080
    ttttaattctg ataagctgca gtcgaatcct ggtgcggcga ccgttggtgt aaactatact 1140
    ccgattcctc tggtagcat ggggtagcat taccgtcatg gtacgggtta tgaaaatgat 1200
    ctcccttact caatgcagtt ccgttatcag ttgataaat cgtgggtctc gcaaattgaa 1260
    ccacagtatg ttaacgagtt aagaacatta tcaggcagcc gttacgatct ggttcagcgt 1320
    aataacaata ttattctgga gtacaagaag caggatattc tttctctgaa tattccgat 1380
    gatattatg tgactgaaca cagtacgcag agatttcagt tgatcggtta gagcaaatat 1440
    ggtctggatc gtatcgtctg ggatgatagt gcattacgca gtcagggcgg tcagattcag 1500
    catagcggaa gccaaagcgc acaagactac caggctatct tgcctgctta tgtgcaaggt 1560
    ggcagcaata tttataaagt gacggctcgc gcctatgacc gtaatggcaa tagctctaac 1620
    aatgtacagc ttactattac cgttctgtcg aatgggtcaa ttgtcgacca ggttgggtta 1680
    acggacttta cggcggataa gacttcggct aaagcggata acgccgatac cattaactat 1740
    accgcgacgg tgaaaaagaa tggggtagct caggctaatt tccctgtttc atttaatat 1800
    gtttccaggaa ctgcaactct tggggcaaat agtgccaaa cggatgctaa cggtaaggca 1860
    accgtaacgt tgaagtcgag tacgccagga caggtcgtcg tgtctgctaa aaccgcggag 1920
    atgacttcag cacttaatgc cagtgcggtt atattttttg atcaaaccac gccagcatt 1980
    actgagatta aggtcgataa gacaactgca gtagcaaatg gtaaggatgc tattaatat 2040
    actgtaaaaa ttatgaaaaa cggtcagcca gttataatc aatccgttac attctcaaca 2100
    aactttggga tggtcaacgg taagtctcaa acgcaagcaa ccacgggaaa tgatggctgt 2160
    gcgacgataa cactaacttc cagttccgcc ggtaaagcga ctgtagtgac gacagtcagt 2220
    gatggggctg aggttaaagc gactgaggtc actttttttg atgaactgaa aattgacaac 2280
    aaggttgata ttattggtaa caatgtcaga ggcgagttgc ctaatatatt gctgcaatat 2340
    ggtcagttta aactgaaaagc aagcgggtgt gatggtacat attcatggtt ttcagaaaat 2400
    accagtatcg cgactgtcga tgcacaggg aaagtcactt tgaatggtaa aggcagtgct 2460
    gtaattaaag ccacatctgg tgataagcaa acagtaagtt aactataaaa agcaccgtcg 2520
    tatatgataa aagtggataa gcaagcctat tatgctgatg ctatgtccat ttgcaaaaat 2580
    ttattaccat ccacacagac ggtattgtca gatatttatg actcatgggg ggctgcaaat 2640
    aatatagccc attatagttc tatgaactca ataactgctt ggattaaaca gacatctagt 2700
    gagcagcgtt ctggagtatc aagcacttat aacctaataa caaaaacccc tcttcctggg 2760
    gttaatgtta atactccaaa tgtctatgcg gtttgtgtag aataa 2805
    <212> Type : DNA
    <211> Length : 2805
        SequenceName : SEQ ID 394
        SequenceDescription :
65
    Sequence
```

```

-----
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
5 atgttagtac ttagcgaaag cttcaagaat aaattgcttc ccatgaatgg gtatatgaaa 60
  ggccggcagcg actccggatc taaagcccag gcacgcgcaa ctgaaaaggg catcgaaactg 120
  cagcgtgaaa tgtggcagac gaacatgcaa aacottgcac cgttcacgcc actcgcctcag 180
  cagtacgtat cacagttgca gaatctttcc tctcttcagg ggcaagggtca ggcgcttaac 240
  cagtattaca actctcagca gtataaagac cttgcagggc aggcgcgcta tcagagtctg 300
  gcagcagcag aggcaacggg tggattaggg totacagcaa caggaaacca gttagcagca 360
10 atcgcaccta cactcgggtca aaactggctg tcagggtcaga tgaacaacta caacaatctg 420
  gcaaatatcg gccttgggtg tcttacaggt caggcaaacg ccggacagaa ctacgctaac 480
  aatgtcagcc aattgtatca acagcaggcg gcagcatcgg cagcaaatgc gaataagcct 540
  tcaggcctac agagttttgc tacaggtgcc attggtgggg ccgcatcagg tgcaatgatt 600
  ggtagtgacg ttcctgttat tgggactggg attggtgctc ttgctggcgg tgttatcggg 660
15 ggtcttggat cattgtttta a 681
<212> Type : DNA
<211> Length : 681
      SequenceName : SEQ ID 395
      SequenceDescription :
20
Sequence
-----
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
25 atgaaaaaaa tattatcagg gttgattctg ctgctttgct gtccttatgg tttcgccgct 60
  aatgggtgatg gcgcaacgca catgtcaaat ttatcatttg gtccgctgac ggtggcagcg 120
  gcgaataacc actccggata caatattttc gaggcactga gcaacacgac tggaaacatac 180
  ccggtgcgct gtcactgtga tgacacgcat ggccggaccgg gccacaacaac agcatttttt 240
  cctatcttct acacgggaga tgccgcaccg gggtttgtgc ttgagcgac tcttaattggg 300
30 ttaaaattact atgctctgaa tgattattta tcggctcggcg tgacgatttt tattattaat 360
  aaccaatatg cggccattcc ttttgaacac ttatccaacc aatccacctc accgcaacat 420
  acctgtggag caggtataata tgggagcact gtaaattctgg attcaggggcg ctccggcaaaa 480
  ttatctttct atgttcggca ttctattact ggcacgggtga caatacccac aacggaagtc 540
  gcttggttgt acgcgggcat gtccgatcat tttoceaaaa cgacccccgt ttctaagggtg 600
35 acaattcgcg gacaactaac ggccccgcag aactgtgagt taacgcctaa tcagagcatc 660
  gatgtcgatt ttcaaaaaat taatagcgct gatgtctcct caacggcggg ttcaattatt 720
  gcagaagaa agattaaaac cgaagtcaca gtatcctgta ccgggatgga agacgtaagg 780
  tccacggagg tggtagtgac gtcgatgatt gcggcaaaaca gaagtgcga tgccaccatg 840
  atcgtgacga gtaatccgga tgtgggaatt aagatttttg ataagaacga ccgtccagtg 900
40 aatgtgggatg ggggcaactt acccgctgat atgggtgcta ttagtcgatt aggaaaaacc 960
  gatggtagcg taacgtttta ttacgcgcc gccagctga cgggcgcaaa accagcgct 1020
  gataatggat ttaccgctac agccacgctg gttattgaat ttactaacta a 1071

<212> Type : DNA
45 <211> Length : 1071
      SequenceName : SEQ ID 396
      SequenceDescription :
50
Sequence
-----
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
  atgaataaaa tatatcggct aaagtggaa aggtcccgt actgttggag cgtctgctcg 60
  gagctgggga gcagagtaaa aggaaaaaag tcccgggctg ttttaattag cgcgataagt 120
55 ttatattcat ctctggtatt cgccgatgat gtcacgtaa accaggataa aactattgat 180
  tttggcaaag agaaccagag catcgattac cgtattacgg tgacagacaa tgccaatctg 240
  gtaatcaatg cgacagatac ttcccgctcg cgtctgactc togtttctgg tgggtgggtg 300
  gatattaccg gaggaagggt aacaatcaat ggcccactta actttttgct gaaagggtacg 360
  ggggttctga atgtctccaa tgctggcagc gagttatatg ctgatgattt gtatgaatca 420
60 aactcaggca tgagacacga tcgcggctat ttaattgtct ccaacggcgg caaaatccat 480
  gttaagggca ccagccgtct gacctaattg cagggaatg tcagtgggtga aggtagccag 540
  gtaaatccg aaaccttctt tatggcggtt tacggcagtt acgggtgtaa tcagtacctg 600
  tcagttaata acggcgggtga agttaatgcc aggaagcaaa ttagcctggg ctattatgat 660
  caagtctccg atacaacact tgctgtttcg gaaggtggta aaatttctgc gcctactatt 720
65 agtttaagca ccaactctga gttagcgtta ggggcacagg aaggaagcgc agcgaaggca 780
  gcagggatta ttgatccga aaaaattgag tttgtgtggg caaagacato cgagaagaaa 840
  atcaccttaa accacacgga taaagacgag actatttccg cggtatttgt cagtggcagc 900

```

```

gagggcctgg gctatatcaa tgcgctcaat ggcacgactt acttaaccgg tgataactct 960
gcctttagtg gtaaaagtcaa aattgagcaa aatggcgctt tagggatcac ccaaaatata 1020
ggtacagcag agatcaacaa ccgcgggaaa ttacacctga aggctgacga tagcatgacc 1080
tttgccaata agatctctgg caacggtaca ataagtatcg acagtgggac ggtggagttg 1140
5 accggcaata actatgcatt cagcggatat attgatgttg cttctggtgc tgtcgctgtt 1200
atttctgaag acaagaatat cggctgtgca gagctggatg tcgatggcaa attgcaaatt 1260
aatgccaaaca aagattgggt atttgataac gatcttgaag gtagaggcat tgttgaaata 1320
aacatggggga atcacgaatt ctcccttcgat gagtttgctt atacagactg gttccagggt 1380
tcactggcgt tcacagaacac gacatttaat ctggaaaaga atgctgagtt tctgcagaaa 1440
10 ggcggggatca ctgcgggtca ggggaagcctg gtaacagtgg gtaagggcgc tcaactccatt 1500
agcacttttg gattctccgg cggaaccgtt gattttggtg ccctgacagc aggtgcacag 1560
atgacagaag ggacgggtcaa cgttagtaaa acgctggatt tgcgcggcga ggggtgtgatt 1620
caggtttctg acagtgcgt tgtccgctcg gtatctcggt atattgactc tgcgttatcg 1680
ctcactgaag tcgatgtgg taacagcacc attaagttgg ttgatgcgca aggtgcggaa 1740
15 gttctggggc atgcggggcaa tctgcaattg caggataaaa atgggcaaat cctctccagc 1800
agcgcccaac gtgatattca gcagaatggg caaaaagcgg ccgtcggcac ttacgactat 1860
cgtctgacga gtgggggtaaa caatgacggt ctgtatattg gttacggcct gaccagctt 1920
gatttacacg ctaccgacag cgatgctctg gtgctgagct ctaacggtaa aagcgagaat 1980
gccgcgcgatc tcagcgcaaaa gattaccggc agtgggtgacc tggcattcag cagccagaag 2040
20 ggtcagaccg tatcgctttc taacaaagac aacgactata ccggtgttac cgatctgcgc 2100
agtgggagata ttttgtttaa taacgataac gtgttgggta ataccatga actgcgtctg 2160
gcggcagaga ctgaactgga catgaatggt cacagccaga ccgtgggcac gctcaatggc 2220
agcgccgatt cactgctgag cttaaatggc ggcagctctga ccgttaccaa cgggggcact 2280
tcaaccgggt cgtaaaccgg gagcggagag ctgaatattc agggcggcac gctggacatc 2340
25 gcgggcgata acagcaacct gacggcgaat ctgaacattg ctaattcggc taatgtcctg 2400
gtaagtcatg cgcagggatt gggtagcgca aacgttgaga acaacggtac cctggcgttg 2460
aataatagcg ctgaaaaaag agcggctgcg tctgtgaatt acgcctcggg cggaatctg 2520
accaacaacg gtacgctgat gaccggaatg tcaggacagc aagctggcaa tgtgttagtg 2580
gtgaagggga actaccacgg taataacggt caactagtaa tgaatacggg actgaatggc 2640
30 gatgactcag taaccgataa attggttgtc gagggcgata ctagcggcac gactgccgtt 2700
acggtgaata acgctggcgg tacaggtgcg aaaaccctta acggtatcga acttatccat 2760
gtagacggta agtctgaggg cgaatttgtt caggctgggc gtatcgttgc gggggcgat 2820
gactacactc tgcgcgtggg acaaggggca actggtatct gaccagcggc 2880
agtgattctc ctgaactgca gccggagcca gaccgatgc cgaatccaga gccaaaccgc 2940
35 aatccagagc cgaaccctaa cccgacacct acgcccgggtc cggatctgaa tgtggataat 3000
gacctgcgac cggaggcggg tagctacatt gcgaaccttg cagcagcgaa taccatgttc 3060
accacagcgtc tgcctgagcg tctgggtaat acgtactata ccgacatggg gacgggtgag 3120
cagaaacaaa ccactatgtg gatgcgccat gaaggtggtc ataataaatg gcgtgatggc 3180
agcggccagc tgaaaaccca aagcaatcgc tatgttctgc aactgggagg cgatgtcgcg 3240
40 cagtggagcc aaaacggcag cgaccgctgg catgttgggg tcatggcggg atatggcaac 3300
agcgacagca aaaccatttc ctgcgcaacc ggttatcgtg caaaagcgag tgtgaacgga 3360
tatagcacag gcctctatgc cacctggtat gccgatgacg agtcgcgtaa tggcgcgtat 3420
ctcgacagtt gggcgcgagta cagctggttt gataacacag tgaaagggga tgacttacia 3480
agtgaatcct ataaatcaaa aggatttacc gcttcaactg aagctggata caaacacaaa 3540
45 ttactgaat ttaatggcag ccagggaacg cgtaatgaat ggatgttca gccgcaagca 3600
caggttacct ggatgggagt caaagccgat aagcaccgcg aaagcaacgg aacctcgtt 3660
catagcaacg gtgatggcaa tgttcaaacc cgaactggcg taaaaacctg gctgaagagc 3720
caccataaaa tggatgacgg taatccccgc gagttccagc cgtttgtaga agtgaactgg 3780
ctacataaca gtaaggattt cagcaccagt atggatggcg tgtctgtcac tcaggatgga 3840
50 gcccgaaata ttgctgagat aaaaaccggg gtggaaggac agctaaatgc caacctgaat 3900
gtctggggga atgtgggcgt tcaggttgcg gataggggat ataatagcac ctctgcaatg 3960
gttggcatta agtggcaatt ctga 3984

```

<212> Type : DNA

<211> Length : 3984

55 SequenceName : SEQ ID 397

SequenceDescription :

Sequence

60 <213> OrganismName : Escherichia coli O157:H7

<400> PreSequenceString :

```

atgataacga tgaaaaaaag tgtattgacg gcgtttataa ctgtggtatg tgcaacgtcc 60
agcgttatgg ctgctgatga taatgctatc acggatggct cagtaacatt taatggtaaa 120
gttattgctc cagcttgtag cctggtagct gcgacgaaag attccgtggg gactttgcca 180
65 gatgttagtg ccacgaagtt gcaaaccaat ggtcaggttt ctggcgtgca aactgatgtg 240
ccaattgaat taagaattg tgatactacc gtaacaaaaa atgcaacggt cacctttaat 300
ggcactgcgg atactactca gattacagcg tttgctaacc aggcctcatc tgatgctgct 360

```

```

acaacactggt ccctgcaaat gtatatgaat gatggtacaa cggccatcaa gccagacaca 420
gaaaccggga acattttgtt gcaagatgga gatcagacgt tgacttttaa agttgattat 480
atcgctacgg ggaaagcgac ttctggtaat gtgaatgcgg taacaaattt ccatattaac 540
tattatttaa 549
5 <212> Type : DNA
  <211> Length : 549
    SequenceName : SEQ ID 398
    SequenceDescription :

10 Sequence
-----
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
15 atgagtaagt ttgtaaaaac agctattgct gcaacaatgg taatgggtgc gttcgtttct 60
  acttcaacaa tcgcccgtgg caacaatggg acagcacgtt tctacggcac cattgaagat 120
  tccccgtgct ctatcgttcc tgatgatcac aaactggaag ttgatatggg tgacattggt 180
  tcagggatcc tgaaaaataa cgggacttct acaccgaaag ctttcagat ccatctgcaa 240
  gactgtgtgt ttgacaccca gacaacgatg accactacct tcaccggtaa cgcgtcttct 300
  accaacacgc gcaattacta caccatttac aataccgata ctggtgcggc atttaacaat 360
  gtcagcctgg ccattggtga cgctcaggga acctcttata aaagcggcgc gggatcga 420
  cagaaaatcg taaacgatac ggcgaccaac aaaggcaag cgaagcagac gctggacttt 480
  aaagcctggc tgggtggcgc tgctgatgcg ccagatttag gtaattttga agccaacacg 540
  acctccaga ttacttatct ctaa 564
25 <212> Type : DNA
  <211> Length : 564
    SequenceName : SEQ ID 399
    SequenceDescription :

30 Sequence
-----
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
  atcggggtta tctttctacg caaggagtat ttatctctac tcccgtcaat gattgcatct 60
  ctttctctctg ctaacggtgt cgcggcgccc attgatttat gccagggata tgatatcaaa 120
35 gcgagttgtc acgccagcag gcaaagcctt tcaggcatta cgcaggctcg gattattgcc 180
  gatgggcaat ggctgggttt ttccgatatg accaataatg ccagcgggtg ggcggtatct 240
  ttgcaacaag gagcggaatt tacattatca ccagaaaaatg aaactggaat gactctgttt 300
  gccataaaca cgttttcagg agaataaat aacggcgggg caatatttgc taaagaaaac 360
  tcaacgctga atcttacgga tgttatTTTT tctggtaacg tcgcaggcgg ctatgggtggc 420
  gcaatctatt cttctgttac taacgatacc ggtgccatcg atttacgtgt cactaacgcc 480
  gtgtttcgca ataactcgc taatgacggc aaagggtggt caatttatac catcaataat 540
  gatatctatt taagtgatga tgtttttaac acgaaccagg catatacatc aacaagttac 600
  agtgatggcg atggcgggcg aatcgatgtc acagataata atagcgacag caagcatcct 660
  tcagggtata cgataataaa taactactgc tttaaaaaa acactgccga aggttatggc 720
45 gggcgcatat ataccaatag cgcgacggct ccctatctta ttgatatttc tgttgatgac 780
  agctacagcc agaacggagg cgtgttagtc gatgagaaca atagcgacg aggtatgga 840
  gatggtcctt cctctgcggc ggggtggctt atgtatctcg gcttaagtga agttaccttt 900
  gatattgccg acggaaaaaac gctggttatt ggcaatacag agaatgacgg agctgttgac 960
  tctattgctg gtaccgggtt aatcaccaaa acagggttccg gcgatctggt acttaatgca 1020
50 gataacaatg actttactgg tgagatgcag attgaaaacg gtgaagttag cctgggccgc 1080
  agcaactccc tgatgaatgt cggcgatacg cattgccagg acgatccgca agactgctac 1140
  ggtctgacga tagggagtat tgataagtac cagaatcagg cagagctaaa tgttggtctc 1200
  acccaacaaa cctttgcgca ctcatgacg ggcctttcaga atggcacttt aaatatcgat 1260
  gctggtggca atgttactgt taatcaaggc agttttgctg gcaccatcga aggtgctggt 1320
55 cagctcacca ttgcgcaaaa cggcagctat gtgctggcgg gggcgcgatc gatggcgcta 1380
  accggcgata tagtgggtgga tgctggtgcg gtgctttcgc tgggaaggcg cgcggcagat 1440
  cttgcccgtc tccaggacga tccgcagtcg atcgtgttaa acggcggtat gctcgatctc 1500
  tctgatttct ccacctggca gagcggtaca tcatacaaa atggccttga agtcagtggc 1560
  agcagcggaa cggttatcgg cagtcaggat gtggtagatc ttgcaggcgg aaacgatatg 1620
60 catatcgggc ggcagcggga agatggcgtc tacgtggtga tcgatgcggg tgacggcgag 1680
  gtcagcctgg caaatgacaa tcaatacctc gccacaacgc aaatcgcttc cgttacgctg 1740
  atggtgagcg acaactcgca gcttggtatg acccattata accgcagggt tatctttacc 1800
  gataagccac aagaaagcgt gatggagatt actgccaatg tcgatactcg ctctacaacg 1860
  actgagcatg ggcgtgatat tgaaatgcgc gccagcgggt aagtggcagt tgatcggggg 1920
65 gtagacacgc agtggggcgc actgatggct gacagcagcg ggcagcatca ggatgagggg 1980
  agcacattga ctaaaacggg ggcgggtaca ctggagctga ccgcagcgg tacaacgcag 2040
  tcagcggtag gagtagaaga gggcacgctg caaggatgat ttgcggatat cttcccttat 2100

```

5 gcttcgctcgc tatgggtcgg tgacgggggca acggttcgtta ctggcgcgga tcaggatatt 2160
 cagtcgaattg atgctacttc cagcgggcact atcgacatca gcgatgggtac ggttttgcgc 2220
 ctgaccggggc aggatacttc cgtcgccctt aatgcctcac tgtttaactg cgatgggacg 2280
 ctggtgaatg ccaccgatgg tgtgacgttg acaggtgagc ttaataccaa ccttgaaact 2340
 5 gacagcctga cttatctttc caacgtgacg gttaatggca atctgaccaa tacgtccggt 2400
 gcggttagcc tgcaaaatgg cgtcgctggc gatacgctga cggtaaacgg tgattatacc 2460
 ggcggcggtg cgctactgct cgatagcgaa ttaaacggcg atgactcggc aagcgatcaa 2520
 ttggtgatga acggtaatac tgctggcaac acaactgtgg ttggttaactc cattacaggg 2580
 attggtgagc cgacatcgac aggcattaaa gtggttgatt tcgcagctga toccacgcag 2640
 10 tttcaaaaaca atgcgcagtt cagtcctggca ggcagcggct acgtcaatat gggagcgtat 2700
 gactacacgc tgggtggaaga taacaacgac ttggtatctgc gatcgcaaga agtaacgcgc 2760
 ccacgcggc acgatccaga cccgactccc gatcctgatc ccacgcagga toctgatcca 2820
 acaccgcgac cgggaacctac gcctgcttac cagccgggtg tgaatgccaa agttggcggc 2880
 tatctcaata acctgcgggc ggcaaatcag gcgtttatga tggagcgacg cgatcacgca 2940
 15 ggtggcgatg gtcagacgct gaatttacgt gttatcggcg gagattatca ttacacagca 3000
 gcggggcaac tgggtcaaca tgaagacact tctacgggtc aacttagcgg cgatctgttt 3060
 agcggggcgt ggggcacgga tggcgagtgg atgcttggga ttggttggtg ctacagcgat 3120
 aaccagggcg acagcgcctc gagtatgacc ggaactcggc ccgataacca gaaccacggt 3180
 tatgcggttg ggtgacctc aagctggttt cagcacggta agcagaagca gggggcctgg 3240
 20 ctggataact ggttgagta cgcggtggtt agcaatgatg tttctgaaca tgaagatggc 3300
 gtggatcatt accattcgtc ggggattatc gcctcgctgg aagcgggta tcagtgggtt 3360
 ccggggcggt gtgtgggtg tgaaccgcag gcgcaggtga tttatcaggg cgtgcagcag 3420
 gatgatttta ccgcgcgtaa ccgtgcgcgc gtgtcacaat cgcaggtgta tgatattcag 3480
 acgcggctgg gtttacacag cgaatggcgt accgctgttc atgtcatacc aacattagat 3540
 25 ctgaattatt atcacgatcc ccattcgacg gaaattgaag aagatgccag cactatcagt 3600
 gacgatgcgg tgaagcaacg gggtgaaacg aaagtgggag tcacgggcaa tatcagtcag 3660
 cgagtttcgc tgcgtggtg cgtggcgtgg cagaaaggga gtgatgattt tgcccagacg 3720
 gcagggtttt tgcgatgac ggtgaaatgg taa 3753
 <212> Type : DNA
 30 <211> Length : 3753
 SequenceName : SEQ ID 400
 SequenceDescription :

Sequence
 35 -----
 <213> OrganismName : Escherichia coli O157:H7
 <400> PreSequenceString :
 atgcactcct ggaaaaagaa acttatagta tcacaattag cattggcctg cactctggct 60
 atcacctctc aggctaattg agcgaccaac gatatttctg gtcaaaactta caatactttc 120
 40 catcactaca acgacgccac ctatgctgat gacgtttact atgatgggtta ttaggcttgg 180
 aacaactatg ccgctgatag ctattacaac ggcgatatct acccggtcat taataacgct 240
 accgttaacg gcgtgatttc tacctactat ctggacgacg gtatttctac caataccaac 300
 gccaatagtc tgacaatcaa aaacagcact attcacggta tgattacctc tgagtgcag 360
 actactgatt gtgctgatga ccgtgctact gggttatggtt atgatcgtct gacactgagc 420
 45 gttgataaatt caacgatcga tgacaactac gagcattata cttacaacgg tacctataat 480
 aatgccgctg acactcatgt tgtagatggt tacgatattg gtactgctat tacactggat 540
 caggaagttg atctgtocat cactaataac tctcatgtag cagggtattac gctgactcag 600
 gggttatgagt ggggaagatat tgacgacaac acagtcagca ctggcgtaaa cagcagcgaa 660
 gtgtttaata acactattac tggttaaagat tctactgtga cctctgggtc atggactgat 720
 50 gaaggtaacta ctggttggtt tggccatact ggtaatgcca gcaactatag caacacgctg 780
 actgcagacg atgttgcaat tgccgcaatc gcaaatccgt atgctgataa tgcgatgcag 840
 actacagtaa ctttagacaa ctcaacacgtg atgggtgatg ttgttttctc cagtaatttc 900
 gatgaaaact tcttccgcga aggtgctaac agctatcgcg atgctgatgg tgatgtagat 960
 accaacgggtt gggatggcac agaccgtatg gatgtgactc tgaacaacgg cagcaagtgg 1020
 55 gttggcgctg caatgtctgt tcatatgggt gatgaagatg gtgatgggtc ttacgacgga 1080
 tatgctgttg gtactgaagc aactgcaact ctgctcgata ttgcagctaa cagcctgtgg 1140
 ccttcatcaa ctgtcgggtg tgataacatc aatactcaat atgacgaaaa tggccatatc 1200
 gtaggaaaacg aagtttaacca gagcgggttg tttaatgtga ctttgaacgg ttggttcagag 1260
 tgggatacaa caaaatcttc tctgattgat actttaagta ttaacagcgg ttcccaagtt 1320
 60 aatgttgagc actctcgtct gatctctgac actgtctctc tgactggcgg ttctaacctg 1380
 aacatcgggtg aagacgggtc tgtagcgact aataccctga ccacgcgaaa tagtaccggt 1440
 aaaatgtctg atgatgtttc tgcgggctgg ggtttagaag atgctgcact gtacgcaaat 1500
 accatcacgc taactaacga cggctcgttg gatattaacg ttgatcagtt cgatgctaac 1560
 ccgttccagg ccgataccct gaatctgacc agtaccactg atactaacgg caacattcac 1620
 65 gctggtgtat tgcatatcca tagcagtgat tacgtaatgg ataccgatct ggtcaacgat 1680
 cgtaccaacg atactaccaa gtcaaacatc aacgataacg taatcgcaat gaactctgat 1740
 ggtcacctga ctattaacgg taacggcgat aacgacaaca ctgcttctat cgaagctggc 1800

```
5  cagaacgaag ttgataacaa cgggtgaccat gttgcagccg cgaccgggtaa ctacaaagtt 1860
    cgtatcgaca acgctactgg tgctgggttct atcgctgact acaacgggcaa cgagctgac 1920
    tacgtcaacg acaaaaacag caacgcgacc ttctctgctg ctaacaaagc tgacctgggt 1980
    gcatacacct atcaggctga acagcgcggt aacaccgttg ttctgcaaca gatggagtgt 2040
10 accgactacg ctaacatggc gctgagcatc ccatctgcga acaccaatat ctggaacctg 2100
    gaacaagaca cggttgttac tcgtttgacc aactctcgtc atggcctggc tgataacggc 2160
    ggcgcgatgg taagctactt cgggtggtaac ttcaacggcg acaacggcac catcaactat 2220
    gatcaggatg ttaacggcat catggtcggt gttgatacca aaattgacgg taacaacgct 2280
    aagtggatcg tcgggtcggc tgcaagcttc gctaaagggt acatgaatga ccgttctggt 2340
15 caggtggatc aagacagcca gactgcctac atctactctt ctgctcactt cgogaacaac 2400
    gtctttgttg atggtagctt gagctactct cacttcaaca acgacctgtc tgcaaccatg 2460
    agcaacggta cttacgttga cggtagcacc aactccgacg cttggggcctt cggtttgaaa 2520
    gccggttacg acttcaaact ggggtgatgt ggttacgtga ctccctacgg cagcatttct 2580
    ggctgtgttc agtctgtgta tgactaccag ctgagcaacg acatgaaagt tgacgggtcag 2640
20 tcttacgaca gcatgcgtta tgaactgggt gtagatgcag gttatacctt cacctacagc 2700
    gaagatcagg ctctgactcc gtacttcaaa ctgggttacg tctacgacga ctctaacaac 2760
    gataacgatg tgaacgggtga ttccatcgat aacgggtactg aagggtctgc ggtacgtggt 2820
    ggctgtgggt ctcagttcag cttaccaacg cctataccga tgctaactac 2880
    ctcggtgggt gtgacgtaga tcaagactgg tccgcgaacg tgggtgttaa atatacctgg 2940
    taa 2943
    <212> Type : DNA
    <211> Length : 2943
        SequenceName : SEQ ID 401
        SequenceDescription :
25
    Sequence
    -----
    <213> OrganismName : Escherichia coli O157:H7
    <400> PreSequenceString :
30 atgaaactca aacatggttg tatgattgtc gtttctgtgt tggcgatgtc gtctgctgcg 60
    gtaagcgcag ccgagggtga tgaatcagta acgaccactg ttaatggcgg tggtattcat 120
    tttaaagggt aagtggtaaa tgccgcttgt gcgattgatt ccgaatcaat gaaccaaacg 180
    gttgagctgg gtcagggttcg ttcttctcgc ctgggtaaag cgggtgacct cagctccgcc 240
    gttggcttca atatcaagct gaatgattgt gataccaatg tttccagtaa tgcagctggt 300
35 gcatttcttg gtactactgt caccagtaat gacgatacgt tagcgtgca gagttcagcg 360
    gcaggctctg cccaaaatgt cggatttcaa attttggacc gtacgggtga ggtattaata 420
    cttgatgggg ccacttttag tgctaaaacc gacttgattg atggcacgaa tatactacca 480
    ttccaggctc gttatattgc tctcgggcag tccgtagctg gtactgcaaa cgcagatgcg 540
    accttcaaag ttcaatatct ataa 564
40 <212> Type : DNA
    <211> Length : 564
        SequenceName : SEQ ID 402
        SequenceDescription :
45
    Sequence
    -----
    <213> OrganismName : Escherichia coli O157:H7
    <400> PreSequenceString :
50 atgaaacttt taaaagtagc agcaattgca gcaatcgat tctccggtag cgctctggca 60
    ggtgttggtc ctcagtacgg cggcgggtggc ggtaaccacg gtgggtggcg taataacagc 120
    ggcccgaatt cagagctgaa tatttatcag tacgggtggg gtaactctgc acttgctctg 180
    caagctgatg ctcgtaactc tgatcttact attaccacg atgggtgggt taacgggtgca 240
    gatgttggtc agggctcaga tgacagctca atcgatctga cccaacgtgg ctttggtaac 300
    agcggcactc ttgatcagtg gaacggtaaa gactctcata tgacagttaa acaattcggt 360
55 ggcggcaacg gtgcagcggg tgaccagact gcatctaatt ccaccgtcaa cgtaactcag 420
    gttggctttg gtaacaacgc gaccgctcat cagtactaa 459
    <212> Type : DNA
    <211> Length : 459
        SequenceName : SEQ ID 403
        SequenceDescription :
60
    Sequence
    -----
    <213> OrganismName : Escherichia coli O157:H7
    <400> PreSequenceString :
65 atgcctattg gtaatcttgg tcataatccc aatgtgaata attcaattcc tcctgcacct 60
    ccattacctt cacaaaccga cgggtgcaggg gggcgtgggt agctcattaa ctctacgggg 120
```



```

ccgttgggat  ctctgtgcgt  atttacgcct  gtaaggaatt  ctatggctga  ttctggcgac  180
aatcgtgcca  gtgatgttcc  tggacttcct  gtaaatccga  tgcgcctggc  ggctctgag  240
ataaactgga  atgatggatt  tgaagtctct  catgatcatg  gtccgctcga  tactcttaac  300
aggcagattg  gctcttcggt  atttcgagtt  gaaactcagg  aagatggtaa  acatattgct  360
5  gtcggtcaga  ggaatggtgt  tgagacctct  gttgttttaa  gtgatcaaga  gtacgctcgc  420
ttgcagtcca  ttgatccctga  aggtaaagac  aaatttgat  ttactggagg  ccgtggtggt  480
gctgggcatg  ctatggtcac  cgttgcttca  gatatcacgg  aagcccgcca  aaggatactg  540
gagctggttag  agcccaaagg  gaccggggag  tccaaagggt  ctggggagtc  aaaaggcgtt  600
ggggagttga  gggagtcaaa  tagcgggtgc  gaaaacacca  cagaaactca  gacctcaacc  660
10  tcaacttcca  gccttcgttc  agatccctaa  ctttggttgg  cgttggggac  tgttgctaca  720
gggtctgatag  ggttggcgcc  gacgggtatt  gtacaggcgc  ttgcattgac  gccggagccg  780
gatagcccaa  ccacgaccga  ccctgatgca  gctgcaagt  caactgaaac  tgcgacaaga  840
gatcagttaa  cgaaagaagc  gttccagaac  ccagataatc  aaaaagttaa  tatcgatgag  900
ctcagcttcc  gccttcgttc  aggggtattg  aaagattgat  ttgttgcgaa  tatagaagag  960
15  caggctaaag  cagcaggcga  agaggccaaa  cagcaagcca  ttgaaaataa  tgctcaggcg  1020
caaaaaaaat  atgatgaaca  acaagctaaa  cgccaggagg  agctgaaagt  ttcacggggg  1080
gctggctacg  gtcttagtgg  cgcattgatt  cttggtgggg  gaattggtgt  tgccgtcacc  1140
gctgcgttcc  atcgaaaaaa  tcagccggta  gaacaaacaa  caacaacaac  tactacaact  1200
acaactacaa  gcgcacgtac  ggtagagaat  aagcctgcaa  ataatacacc  tgcacagggc  1260
20  aatgtagata  cccctgggtc  agaagatacc  atggagagca  gacgtagctc  gatggctagc  1320
acctcgctga  ctttctttga  cacttcacgc  atagggaccg  tgcagaatcc  gtatgctgat  1380
gtttaaactc  cgctgcagtg  ttgcagggtg  ccgacttcta  attctaatac  gtctgttcag  1440
aatatgggga  atacagattc  tgttgatat  agcaccattc  aacatcctcc  ccgggatact  1500
actgataacg  gcgcacggtt  attaggaaat  ccaagtgcgg  ggattcaaag  cacttatgcg  1560
25  cgtctggcgc  taagtgtgtg  attacgccat  gacatgggag  gattaacggg  ggggagtaat  1620
agcgtctgtg  atacttcgaa  taaccacca  gcgccgggat  cccatcgttt  cgtctaa  1677

```

<212> Type : DNA

<211> Length : 1677

30 SequenceName : SEQ ID 404
SequenceDescription :

Sequence

35 <213> OrganismName : Escherichia coli O157:H7

<400> PreSequenceString :

```

atgttttcta  ctttcaaaaa  agcagctctg  ctggcagcta  ttgcattacc  tttttcaact  60
atggctgcgc  ctacagtcac  ttttcagggg  gaagtaaccg  atcagacctg  ttccgtaaat  120
atcaacggtc  aaaccaattc  agtagtattg  atgccgaccg  tagccatggc  tgacttcggg  180
40  gcaactttag  ctgatgggtc  gagcgaggcc  cagacgccgt  ttacggtttc  tgtgtctaac  240
tgccaggctc  caactgggtc  agatcaggca  atcaacacca  ccttcctggg  ctacgacgtt  300
gacgctagca  cgggtgttat  gggaaaccgt  gataccagca  gcgatgcggc  gaaaggcttt  360
ggcattcagt  taatggattc  cagcacttct  ggtaaccag  taactctggc  tggcgcgact  420
aacgtaccgg  gtctgacct  gaaagtgggc  gataccgaag  ccagctacga  cttcgggtgcg  480
45  cgttacttcg  ttatcgatag  cgctgctgcc  actgccggta  aaattaccgc  tgtcgcagaa  540
tacacctga  gctacctcta  a

```

<212> Type : DNA

<211> Length : 561

50 SequenceName : SEQ ID 405
SequenceDescription :

Sequence

55 <213> OrganismName : Escherichia coli O157:H7

<400> PreSequenceString :

```

atgaacagtg  aaggaggaaa  accggggaat  gtactgaccg  ttaacggcaa  ctataccgga  60
aacaatggcc  tgatgacgtt  caacgcgacg  ctgggcgggc  ataattcacc  caccgataag  120
atgaacgtga  aaggcgatac  ccaagggaac  actcgcgttc  ggggtgataa  cattggcggc  180
gtcgggtcgc  aaacgggtcaa  cgggtattgaa  ctcatagagg  ttggcggtaa  ttctgcaggg  240
60  aatttcgcgc  tgaccaccgg  aactgtcgaa  gctggggcgt  acgtctacac  gctggctaaa  300
gggaagggga  atgacgagaa  aaactgggat  ctgaccagta  aatgggacgg  cgtaacgcca  360
gcggatacac  ccgatcccat  caataatccc  cctgttgttg  atccggaagg  cccatcagtt  420
tatcgcccg  aggcgggaag  ctatatcagc  aacattgccc  cagccaactc  gctgttttagc  480
catcgcttac  acgaccgtct  gggtagccg  caatatacag  attcactgca  ttctcaggat  540
65  tcagcaagca  gtatgtggat  gcgtcatgtc  ggggggcacg  aacgttcacg  tgccggagac  600
ggccagctaa  atactcaggc  taaccgctat  gtattgcagc  taggcggcga  tttggcgacg  660
tggagtagca  acgcgcagga  tcgtggcgat  cttggcggtg  tggcaggcta  cgccaatcac  720

```

```

cacagtaata ctcagagtaa tcgtgtgggt tataaatcgg atgggcccgc atcgcggttac 780
agcgctgggc tgtacgcgac ctgggtatcag aacgatgcga ataagaccgg cgcttatgtt 840
gacagctggg cgctgtataa ctgggttgat aacagcgtca gttccgataa ccgttctgct 900
gacgactatg attctcgcgg tgtgacggcc tctgttgagg gtgggtatac ctttgaagcg 960
5 ggaacatgta gcggcagcga agggacgctg aatacctggg acgtccagcc acaggcgcaa 1020
atcacctgga tgggtgtgaa agattctgac catgcccggg aagacggaac gcgcattgaa 1080
acggaaggcg acggaacgtg gcaaacgcga cttgggggtg aaacctacct gaatagccat 1140
caccagcgtg acgatggtaa acagcgtgag ttccagcctt acattgaagc gaactggatc 1200
aacaatagca aagtctacgc cgtgaagatg aatgggtcaa ccgtaagccg tgatgggtgcg 1260
10 cgaaatctcg gtgaagtacg taccgggggtt gggcggaag taaataacaa ccttagcctg 1320
tgggggaatg tcggtgtgca actaggtgat aaaggctata gcgatactca gggcatgctg 1380
ggagtgaat atagctggta a 1401

```

<212> Type : DNA

<211> Length : 1401

15 SequenceName : SEQ ID 406

SequenceDescription :

Sequence

```

20 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgtcatatc tgaatttaag actttaccag cgaacacac aatgcttgca tattcgtaag 60
catcgttttg ctggtttttt tgtccggctc tttgtgcct gtgcttttgc cgtacaggca 120
cctttgtcat ctgcogaact ctattttaat ccgcgctttt tagcggatga tccccaggct 180
25 gtggccgatt tatcgcgttt tgaaaatggg caagaattac cgccagggac gtatcgcgct 240
gatattctatt tgaataatgg ttatatggca agcgtgatg tcacatttaa tacggcgac 300
agtgaacaag ggattgttcc ctgcctgaca cgcgcgcaac tcgccagtat ggggctgaat 360
acggcttctg tcgcgggtat gaatctgctg gcgatgatg cctgtgtgcc attaacaca 420
atggtccagg acgctactgc gcatttagat gttggtcagc agcactgaa cctgacgatc 480
30 cctcaggcat ttatgagtaa tcgcgcgctg ggttatattc ctctgagtt atgggatccc 540
gggtattaatg ccggattgct caattataat ttcagcgga atagtgtaca gaatcggatt 600
gggggtaaca gccattatgc atattttaac ctacagagtg ggttaaataat tgggtcggtg 660
cgtttacgcg acaataccac ctggagttat aacagtagcg acagatcatc aggtagcaaa 720
aataaatggc agcatatcaa tacctggctt gacgagaca taataccgtt acgttcccgg 780
35 ctgacgctgg gtgatggtaa tactcagggt gatattttcg atggtattaa ctttcgcggc 840
gcacaattgg cctcagatga caatatgtta ccgatagcc aaagaggatt tgccccgggtg 900
atcggcggtg ttgctcgtgg tactgcacag gtcactatta acaaaaatgg gtatgacatt 960
tataatagta cgggtgccgc ggggcctttt accatcaacg atatctatgc cgcaggtaat 1020
agtgtgact tgcaggtaac gattaaagag gctgacggca gcacgcagat ttttaccgta 1080
40 ccctattcgt cagtcggctt tttgcaacgt gaagggcata ctcgttatc cattacggca 1140
gggaaatacc gtagtggaaa tgcgcaacag gaaaaacccc gctttttcca aagtacatta 1200
ctccacggcc ttccagctgg ctggacaata tatgggtgaa cgcaactggc agatcggtat 1260
cgtgctttta attttggtat cgggaaaaat atgggggcac tgggcgctct gtctgtggat 1320
atgactcagg ctaattccac acttcccgat gacagttagc atgacggaca atcgggtcgt 1380
45 tttctctata acaaatcgct caatgagtc ggcacgaata ttcagttagt gggttaccgt 1440
tattcgacca gcggatattt taatttcgct gatacaacat acagtcgaat gaatggctac 1500
aacatcgaaa cacaggacgg agttattcag gtttagccga aattcaccga ctattacaac 1560
ctcgcttata acaaacgcgg gaaattacaa ctaccggtta ctacgcaact cgggcgctca 1620
tcaacactgt atttgagtgg tagccatcaa acttattggg gaacgagtaa tgcgatgag 1680
50 caattccagg ctggattaaa tactgcgttc gaagatatca actggacgct cagctatagc 1740
ctgacgaaaa acgcctggca aaaaggacgt gatcagatgt tagcgcgtaa cgtcaatatt 1800
cctttcagcc actggctgcg ttctgacagt aaatctcagt ggcgacatgc cagtgccagc 1860
tacagcatgt cacacgatct caacggtcgg atgaccaatc tggctggtgt atacggtacg 1920
55 ttgctggaag acaacaacct gcctatagc gtgcaaaccc gctatgccgg gggaggcgat 1980
ggtaatatgc gaagcacagg ctacgccacg ctgaattatc gcggtggtta cggcaatgcc 2040
aatatcggtt acagccatag cgatgatatt aagcagctct attacggagt cagcggtggg 2100
gtactggctc atgccaatgg cgtaacgctg gggcagccgt taaacgatac ggtggtgctt 2160
60 gttaaagcgc ctggcgcaaa agatgcaaaa gtcgaaaacc agacgggggt gcgtaccgac 2220
tggcgcggtt atgccgtgct gccttatgcc actgaatatc gggaaaaatag agtggcgctg 2280
gataccaata ccctggctga taacgtcgat ttagataacg cggctcgctaa cgttggtccc 2340
actcgtgggg cgatcgtgcg agcagagttt aaagcgcgcg tggggataaa actgctcatg 2400
acggttaaccc acaataataa ggcgctgcgg ttggggcgca tggtgacatc agagagtagc 2460
cagagtagcg gcattgttgc ggataatggt caggtttacc tcagcggaat gcctctagcg 2520
ggaaaagttc aggtgaaatg gggagaagag gaaaatgctc attgtgtcgc caattatcaa 2580
65 ctgccaccag agagtcagca gcagttatta acccagctat cagctgaatg tcgttaa 2637

```

<212> Type : DNA

<211> Length : 2637
SequenceName : SEQ ID 407
SequenceDescription :

5 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :

```
10 atgcagataa tctttggaga aaaatgcgtg tcattactac gactatTTTT tgccgcgctc 60
   ttaatgctat ggtgcgctca aaccgctgct tatagcgggc agtgtcatac cactcagggg 120
   aatccgtata ttggcgctcaa ttttggcggt aaaaccctgg aggaagaaga aaatacgact 180
   ggggtagtaa aagacaaatt ttatcagtggt aacgaatcga atgattatta tgtttcctgt 240
   gattgcgata aagacaatgt cagaagtggc cgatgggcat tcgccgcgga ttcaccgtta 300
   gtctattttag gcgacaactg gtacaaaatt aatgactatc ttgccgccaa agttttattg 360
15 cagggttaaag gcagttctcc tacagcggtt cctttcgaaa acgtggggac tggggcagat 420
   acccggtggc atattttgtga ccccgcggtt caacgtttag gcggccaggg agctagcggg 480
   aatagcggta gcttttccct gaaaatattg cagccgttcg ttggttcggg cgtcattcct 540
   cctatggcgc tggcgcgatt atttgaatgc tacaacatac ccgcaggtga ttcctgcacg 600
   actacaggca caccggtttt agtgtattac ctgtctggta ctatcaattc acttggtcga 660
20 tgttcctgca atgccggaga aacaatcgag gtcgatctgg gcgacgtatt tgcggctaac 720
   tttcgtggtg taagggcataa gcctcctggg ccagaaacgg cagaacttgc aattccagtc 780
   aggtgtaaca cgggaaacgc ggggttagtt aacgtcaacc tgagtctgac ggcaaccaca 840
   gacccagct atccccaggc gattaagacg tcacgtcctg gcgtggcggt ggtggtgacc 900
   gatagccaga acaacattat ttccctgctt ggtggaacat taccgctctc tattcctgat 960
25 gatgcagaca gtatcgcggtg a
```

<212> Type : DNA
<211> Length : 981
SequenceName : SEQ ID 408
SequenceDescription :

30 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :

```
35 atgaaaaatta aaactctggc aatcgttggt ctgtcggctc tgtccctcag ttctacagcg 60
   gctctggcgg ctgccacgac ggtaaatggt gggaccggtc actttaaaagg ggaagttytt 120
   aacgcgcgctt gcgcagttga tgcaggctct gttgatcaaa ccgttcagtt aggacagggt 180
   cgtaccgcat cgctggcaca ggacggagca accagttctg ctgtcggttt taacattcag 240
   ctgaatgatt gcgataccaa tgttgcatct aaagccgctg ttgccttttt aggtacgggtg 300
40 attgatgcgg gtcataccaa cgttctgggt ctgcagagtt cagctgcggg tagcgcaaca 360
   aacgttgggt tgcagatcct ggacagaacg ggtgctgcgc tgacgctgga tgggtgcgaca 420
   ttcagtgagc aaacaacctt gaataacggt actaacacca ttccgttcca ggcgcggtat 480
   tatgcaatcg gcgaggcaac cccgggtgct gctaattgcg atgcgacctt caaggttcag 540
   tatcaataa
```

<212> Type : DNA
<211> Length : 549
SequenceName : SEQ ID 409
SequenceDescription :

50 Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :

```
55 atgaaattaa aagtcacgcg tacactgatt gctactgttg ccgtgggtgt aagctttaac 60
   agcaattttg cttctgcgag tacaacgtcc gctcttttaa ccgtaaacag taacctgact 120
   atgggtacct gcagtgctca gataatggat aatagtaata aagtgatcaa tgaagtgggtc 180
   tttggcaatg tttatatttc tgaactcggg gcaaaaagca aagtgaaca gtttaaaatt 240
   cgcttttagca attgctctgg ccttcccaaa aacagcgccc aaatagtgtt ggcacctaat 300
   ggtatattct gtgctggttc tcaatcgtca tcggcgggtt tttctaaca gtttactgac 360
60 gctagcgcag caaccagaac ggctgtggaa gtatggacta cagatacacc ggaaagcaat 420
   ggcagtacgc aatttcattg tgctcaaaag ataccagtgc ctgtgacgct tcccgcgac 480
   accacaactc agccttaccg ttaccggtta atgacagga tgaccgttgc ggaaggtaga 540
   ttggtaacgg atgtaagacc gggtaatttc cgctctccca cgactttcac gatcacttat 600
   cagtaa
```

<212> Type : DNA
<211> Length : 606
SequenceName : SEQ ID 410

SequenceDescription :

Sequence

5 <213> OrganismName : Escherichia coli O157:H7
 <400> PreSequenceString :
 ttggcatcaa cagttgagta tggtagagaca gttgatgggtg ttgtcctgga aaaagatatac 60
 cagctggttt atgggaccgc caataatacag aaaatcaatc ctggcggaga acagcatatt 120
 aaagaatttg gtataagtag taatactgaa attaacggcg ggtatcagta cattgaaatg 180
 10 aatggcaccg cagaatactc agtattaaat gatggttatc aaattgttca aatgggtggc 240
 ggggcaaac agactacgct caataatggt gtgtacagg tttatggcg agcgaatgat 300
 cccacgatta aaggcgggcg cttaatcggt gaaaaagatg ggattaccgt ccttgccgct 360
 atcgaaaagg gaggattact ggaggttaa gaggggggat tagcgattgc ggtagatcag 420
 aaagtcaggc gctgtattaa agcaagcacg cgggtcatgg aggtattcgg aacaaaccgt 480
 15 ctcggtcagt tcgaaatcaa gaatggtatt gctaacaata tgctgttgg aaacggcggga 540
 agtttgcgag ttgaagaaaa tgacttcgct tataatacta ctgtagatag tggcggctta 600
 ctggagggtta tggatggcgg gactgcaact ggcgttgata aaaaagcagg cggaaaatta 660
 attgtctcaa cgaatgcgct ggaagtgaat ggtacaaaca gtaaggcca atttagtata 720
 aaagatgggtg tgtcaaaaaa ttatgaactg gatgatgggt cggggttat tgttatggag 780
 20 gacacgcagg ccattgacac tatcctcgat gacgatgcca ctatgcaatc gctgggaaag 840
 gatactggta cgagagtgc ggcaaatgcg gtatatgatc tcggtcgatc agatcagaat 900
 ggaagtataa cgtattcctc taaagccatc tctgaaaata tggttatcaa caatggcgcg 960
 gctaactctt gggctggcac aatggttaac gtgtcagtca gaggaatga tggcattctt 1020
 gaggttatga agcgcgcaat aaattatgca ccgcgaatgt tgggtgggtaa ggtagtgggt 1080
 25 tctgaggggc ctctcttaag aacgcattgg ggcgtggata ccagcaaagc ggatgtttcg 1140
 ctcgaaaata gcgcattggac catcattgcc gatatacta cgacgaacca aaacaccgcg 1200
 cttaacttag ccaaccttgc gatgtctggc gcaaatgtga ttatgatgga tgagtcagt 1260
 actcgttcat ctgtgacggc aagtgcggaa aatttcaacta cgttgaccac caataccctg 1320
 tcgggaaacg gcaattttta tatgctgacc gatatggcga atcatcagag cgatcagctc 1380
 30 aacgtcaccc gtcaggcaac aggtgatttc aaaatattcg tgacggacac cgggtgccagc 1440
 ccggcagcag gagatagcct tacactggta acaacggcg gcggtgatgc tgcatttacg 1500
 ttgggcaatg ccggaggcgt tgttgatata ggtacgtatg aatatacctt gctggataat 1560
 ggtaaccata gctggagtct ggcagagaat cgcgcgcaaa ttacccttc aaccactgat 1620
 35 gtgctgaata tggcggccgc acaaccgctg gtatttgatg cagaactgga caccgtgcgt 1680
 gagcgtcttg gtacgtaaa aggcgttagt tacgatacgg cgatgtggag ttcggcaatt 1740
 aacaccgcga acaacgtgac cactgatgcg ggagctgggt ttgagcaaac attgacgggc 1800
 ctgacgctcg gtatcgatag ccgtttctcc cgtgaagaaa gcagcacaat tcgcggcttg 1860
 ttctttgggt actctcattc tgatattggt tttgatcgcg gcggcaaaagg caatgtcgat 1920
 agctataccc tgggggctta tgccgggttg gagcatcaga acggtgccta tgttgatgga 1980
 40 gtggtgaaag ttgaccgttt tgccaacacc atccatggca agatgagtaa tggggcaaca 2040
 gcgtttggcg attacaatag taacggcgcg ggtgctcatg tcgagagcgg gttccgttg 2100
 gttgacggat tgtgagtggt tagaccctat ctggccttta cggcctttac cacagatgg 2160
 caggactaca cgttatcaa cggcatgcgc gcggatgtgg gaaatacccg gatattacgc 2220
 gctgaagcgg gaacggcggg aagctatcac atggacctgc aaaacgggtac gacgctggaa 2280
 45 ccctggctga aagccgcgct gcgtcaggaa tacgccgatt ctaaccagg gaaagttaa 2340
 gacgatggca aatttaataa tgatgtggct ggaacccgtg cgttttatca ggctgggata 2400
 aggtcatcgt ttaccccgac gttaaagggt catttgcag tcagctatgg caatggcgca 2460
 ggggtagaat cgccgtggaa taccaggcg ggtgtggtct ggacgttctg a 2511

50 <212> Type : DNA
 <211> Length : 2511
 SequenceName : SEQ ID 411
 SequenceDescription :

55 Sequence

<213> OrganismName : Escherichia coli O157:H7
 <400> PreSequenceString :
 atgcaaagga aaggcaataa actgttgatt cagttatgca gtgtgatact gctatttttt 60
 60 accacatcct ggtatgcatt ggcaaatgaa tgttatatag agagaaatgc tgaaggggat 120
 tatcacatga agataagctc tactcagctt agtctggcgt cacaaatggg cgagggtccg 180
 acagaaatag ccgaagctac atgggatgta aatattcaac taagaggcga tgccataggg 240
 tgtaaatctc ttggggatag taaggcagtt cactttctta atacagctga cccaagttaa 300
 atatccacgt acaccacaac gaatggcgca gcgttatata aaacaactgt tccaggcatt 360
 65 gtgtattctg tcgagttatt atgccttagt tgtggtgccg cagatgaact tgatttatgg 420
 ctacctgcac aaagtggcg agataacttc ataccaagca ccagacgaa atgggcctat 480
 gagtacagt atcaaatgtg gtatttacgt tttcgcttat tcataactcc tgaatttaaa 540

```

    cccaagaatg gtgtttccag cggaacaacg atagcaggaa agattgcgtc atggtatata      600
    ggtaccaatg accagccgtg gatcaacttt tacattgaca atgactcttt aaagtttttc      660
    gtcgatgaac cgacctgtgc aacagttgcc ctggcacaag atcagggcaa cgtcagtggtc      720
    aatcaggtaa cgcttgggaa cagctatgtt tcggaagtga aaaatgggct tacgcgggaa      780
5   atcccttttt ctatccgtgc tgaatactgt tatgccagta aaattacggt taagttgaaa      840
    gcggcaataa aaccagcga tgccacactg gtgggtaaaa cgactggctc ggcttcaggc      900
    gtggctgtaa aagtaaattc aacttatgac aatagcaaag tattgttaaa agcagatggg      960
    agcaacacgg ttgactacaa cttcgccgcc ttggtcaaaa acctgctgtt ttacaccttt     1020
    acggcgacgc tggtagcgga tggtagcggt aatgctgtcg gtgttggaac attttcagggt     1080
10  aacgcgacct tctcctttac ctacgaataa                                     1110
    <212> Type : DNA
    <211> Length : 1110
        SequenceName : SEQ ID 412
        SequenceDescription :

15  Sequence
    -----
    <213> OrganismName : Escherichia coli O157:H7
    <400> PreSequenceString :
20  ttgtaccagt ttactcatca aaaaagccgt atccccgaaa aaacgctact tgcggccctgt      60
    tgtgccctgt tttatagcag caacggtgct gcggcggaca ccgtggaata tgacagttcc      120
    tttttaatgg gaactggcgc atcaacgatt gatgttaaac gttatgctca aggcaacccg      180
    acaccgccgg gtctctataa tgtccgcgta tttgtaaaac gtcaggcgac ttccagctta      240
    gaaattccgt ttgtggatat tggcgaaaac agtcggcgcg cctgtcttac ccataaaaac      300
25  ctggcgcaac ttcacattaa gcaacctgaa cagcctgtca ctttactcgc cagagaagggt      360
    gaagaagagg attgtctgga tctggcgaag tcatacgaag aggcggatgt gtgctttgac      420
    ggtagtgaac agtttctcga tctgacgatc cctcaggcct atgttctgaa aagctatggc      480
    ggctacgttg acccttcttt atgggaatcg ggaattaacg ctgccacact ggcatatacc      540
    ctgaacgcgt atcacacaag ttcagataac gacaatagtg acagcgtcta tggcgcgcttc      600
30  aactcaggta tcaatttagg agcctggcac tttcgtgcgc gcggtaacta taactggaca      660
    acagataacg gcagcgatgt cgatttccag gatcgttact tacagcgtga cattccggca      720
    atccgttccc agataaattat gggtagtccc tataccaccg gtgaaacggt tgactctgtc      780
    aacgtccgtg gtgttcgcct gtacagcgac atcccgtagt gccttcggc gctggccagt      840
    tacgctccga ccatccgcgg tgtagcaaac tccaacgcca aagtcaccgt gacgcaaagc      900
35  ggatataaaa tttatgaaac caccgttccg cccggtgaat ttgttataga cgacattagc      960
    ccttccggct ttggtagcga actgggtcgtg accattgaag aagcggatgg ttccaaacgc      1020
    acctttacgc ttacccttctc gtccggttga gtccgtgcaa gtccgtgtgt gggccgcttg      1080
    gatttcagcg cgggtaaagt cattgatgac agtctgcgat ccgaacccaa tatggggcaa      1140
    gcctcttatt actatggtct gaataacctc ttcacggggt ataccggcat tcagttcacc      1200
40  gataataact atcttgccgg gctgttaggt gtgggtatca acaccagcat cggcgccctt      1260
    gcggtagacg ttaccatttc cctgtctgaa atcccgtagt ataaaacctc ccaggggcaa      1320
    agttatcgcg tgacctggaa caaacttttc caggataacc ggacatcatt taacctcgcg      1380
    gcgtaccgct attccaccca ggattacctg ggcctgcatt atgcgttagt cctcattgac      1440
    gacgccaaagc atttgtctgc cgatgaagac aaaaacacca tgcagacgta ctacagtatg      1500
45  aaaaaccagt ttaccgtcag cattaaccag ccatgtaata tcgcttatga agattacggt      1560
    tcgctgttta tttccggtag ctggacgtat tactgggcgg cgaacaatag ccgcactgaa      1620
    tataatgttg gttacagtaa aagcgttttcg tggggcagtt tcagcgtcaa cctacaacgt      1680
    agctggaatg aagacggcga gaaagatgac gcgatgtacg tcagcgttag cgtacctatt      1740
    gagaatattt taggtggcaa acgtaagctc tctggtttcc gcaatttaaa tactcagctc      1800
50  aataccgatt tcgatgggtc acatcagttg aatgttaaca gttccggtaa cactgaaaac      1860
    aatctggtga actacagtgt caacgcaggt tatagcctcg ataaaaacgc cggcgattta      1920
    gcctctgttg gtggttatct caactatgaa tctgggttag gcggtatttc cgcttcggcc      1980
    tcggccactt ctgataacag ccaacagtac tccatctcaa ccgatggcgg ctttgtatta      2040
    cacagtgttg gtttaacggt cactaacaac agtttcagca gtaacgacac gctgggtgta      2100
55  atcaacgccc taggtgctaa aggcgcacga atcaataaca gtaataacga aatcgatcgc      2160
    tggggatatg ccgtgacgtc ctctgtcagc ccatatcgtg aaaaaccgggt aggtctgaac      2220
    attgaaaacac tggaaaacga tgttgaaactg aaaagtacca gcgccaccac cgtaccacgt      2280
    agcggctccg ttgttttgac ccgtttcgaa actgacgagg ggcgttctgc cgtgctgaat      2340
    attactgccg ccaatggcaa atccattccg tttgctgcgg aggtttacca ggggtgaggtg      2400
60  atgatcgcca gcatgggcca ggggtggctag gcatttgtac gcggtattaa cgacagcggg      2460
    gaattaatcg tgcgtgggta tgaaaacaac caaaccattg actgtaagtt gcactaccag      2520
    ttcccggcgc agccacaaac gcagggaagc accaacacct tattacttaa caatcttacc      2580
    tgtcaggtag caaatcacta a                                     2601
    <212> Type : DNA
    <211> Length : 2601
        SequenceName : SEQ ID 413
        SequenceDescription :

```

Sequence

5 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgaagtcca aacgattgct gcatagcggc atcgccagtt tgagtctggt tgcctgcggg 60
gtgaatgcgg cgacggatct tggcccggca ggggatattc atttctccat cactatcacc 120
actaaagctt gcgagatgga aaaaagcgat ctggaagtcg atatgggaac aatgacgctg 180
caaaaacctg cggcagtcgg tacggtgttg agcaagaaag atttcacat tgaactcaaa 240
10 gagtgcgcatg ggatatccaa agcgaccgtt gagatggaca gtcagtcgga cagcgatgat 300
gattccatgt ttgcccttga ggctggtggc gcaacgggtg ttgctgtgaa gatagaggac 360
gataaaggaa cgacgcaagt tcccaaaggc tccagcggaa cgccgattga atgggcgatt 420
gatggcgaaa ccacgtcgct tccctaccag gcgagttatg tggctcgtcaa cactcaggcc 480
actggtggca cagcgaatgc ccttgtaaat ttttccatca cctatgagta a 531
15
<212> Type : DNA
<211> Length : 531
SequenceName : SEQ ID 414
SequenceDescription :

Sequence

25 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgaaatata ataacattat tttcctcggg ttatgtctgg ggtaaccac ctattctgct 60
ttatccgcag atagcggtat taaaattagc gggcgcgtcc tcgattatgg ctgcacagtc 120
tcacgcgatt cgcttaattt taccgtagat ctccaaaaaa acagtgccag acaatttcca 180
acgaccggta gcacaagtcc agccgtccct tttcagatta cgttaagtga atgcagcaaa 240
gggacaacgg gggttcgggt tgcatttaac ggtattgagg acgcagaaaa taatactctg 300
30 ttgaaactgg atgaggggaa caatacggcc tccggtttag gtatagaaat actggacgga 360
aatatgcgtc cgttgaaact gaatgacctt catgccggga tgcagtggat cccactggta 420
ccagaacaga acaatatattt gccttactcc gctcgtctga agtcaactca gaagtcgctc 480
aatccgggac tggtaggggc ttcggcaacc tttacccttg aatttcaata a 531
35
<212> Type : DNA
<211> Length : 531
SequenceName : SEQ ID 415
SequenceDescription :

Sequence

40 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgaaatggc gcaaacgtgg gtattttatt gcggcaatat tggcgctcgc aagtgcgacg 60
45 atacaggcag ccgatgtcac catcacgggt aacggtaagg tcgtcgccaa accgtgcaca 120
gtttccacca ccaatgccac gggtgatctc ggcatcttt attotttcag tctgatgtct 180
gccggggcgg catcggcctg gcatgatgtt gcgcttgagt tgactaattg tccggtggga 240
acgtcaaggg tcaactgccag ctccagcggg gcagccgaca gtaccggata ttataaaaac 300
caggggacgg cgcaaaacat ccagtttagg ctacaggatg acagtggcaa cacattgaat 360
50 actggcgcaa ccaaaacagt tcaggtggat gattcctcac aatcagcgca cttcccgtta 420
caggtcagag cattgacggt aaatggcgga gccactcagg gaaccattca ggcagtgatt 480
agcatcacct atacctacag ctga 504
<212> Type : DNA
<211> Length : 504
55 SequenceName : SEQ ID 416
SequenceDescription :

Sequence

60 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgaaaagag cgctcttat aacaggactt ttgttgatat ccacatcctg cgcttatgcc 60
tcctcagaag ggtgtggagc tgacagcact agcgggtgca caaattacag cagtgtgggt 120
gatgatgtta cgttgaacca gacagataac gtgacaggac gggagtttac ctctgcaacg 180
65 ctaagtagca ctaactggca atacgcctgt tctgtctctg cgggtaaggc agttaaaact 240
gtctatatgg tcagccccgt acttaccacc actggacatc agacaggata ttacaaactc 300
aatgacagcc tggatatata aaccatgaac cgccccgga atcctggaga cttaa 354

<212> Type : DNA
<211> Length : 354
SequenceName : SEQ ID 417
5 SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
10 <400> PreSequenceString :
atgaaaaaag cacttctcgc agccgctctg gttatggcgt ctggttccgc cctgggtgta 60
gatgggtggtc atatcgactt taacgggtatg gtacagtcgc gtacctgtaa agtgggtgtg 120
gtagatactg gtatgcatag cgttaccact gatggcggtg ttacctgga tactgcgaat 180
gttactgata cttttgctga agttagcgca actgctgtcg gtttactgcc gaaagagttc 240
15 atgattttctg ttgagtgtga tccagggtgct ccgaagaatg ctgagttaac tatgggttct 300
gcaagttacg cgaacaccag cggtagccctg aataacaata tgaacatcac tgttaacggt 360
attgcaccgg ctcagaacgt aaacattgca gttcataaca tgaaaaacaa agctggcgct 420
gctgaaatta agcagggtcca tatgaacaac tcttctgaag ttcaggaact gacattagac 480
gcagaaggta aaggccagta cgtatttaac gcatcttacg ttaaagcacc gaacagcccg 540
20 gctgtaactg ctggtcatgt aaccactaac gcgctgtaca ccgttgctta taagtaa 597

<212> Type : DNA
<211> Length : 597
SequenceName : SEQ ID 418
25 SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
30 <400> PreSequenceString :
atgaaaccaa atatgattgt aggagcatta gcgttaactt ctgtgtttat ggcaggtcac 60
ctacaggcgg ctgatggaac agtccatttc cgtggtgaaa ttattgacag tacttgcaa 120
gtcactcctg aaactaaaga tcaggctcgtt gatttaggca aagtaaaccg tacagccttt 180
agtggcgctg atgatgtggc tgccccgacg gctttttcta tcgatctgac tcaatgcccg 240
35 gaaaccttta agtccgcccgc aattcgtttc gatggtaatg aagatgctca tggtaatggc 300
aacctggcaa ttggtacccc gctggataac tctaacgatg ctgccgctgg tattagcccg 360
agtataaca gtggggatta tactggtgcg ggtgccgtta gtgcagcgaa aggcgtagct 420
attcgtttat ataaccgtgc agataacact caggtaagat tatatgaaaa ttctgcatca 480
actccgattt ctaatggtaa tgcattccatg aagttcatgg ctcgttatat tgctacggaa 540
40 acgactattg accctggtac agctaaccgc gactcgcagt ttacagttga atatataaaa 600
taa 603

<212> Type : DNA
<211> Length : 603
SequenceName : SEQ ID 419
45 SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
50 <400> PreSequenceString :
gtgccaattt tccagcgtga aggccatctc aaatatagct ttgccgcagg tgaatatcag 60
gccgggaatt atgacagcgc ctgccgcgct ttccggcagc ttgatctgat ctacggttta 120
ccgtggggga tgacggccta cggcggcgta ttaatctcta ataattacaa tgcatttaca 180
ttagggatag ggaaaaactt tggttatata gggggcattt ccattgatgt gacgcaggct 240
55 aaaagcgaac tgaataacga tcgcgatagc cagggacaat cttatcgttt cttatatcc 300
aagagcttcg aaagcggcac cgatttccgc cttgcgggct atcggtactc taccagcggg 360
ttctatacct tccaggaagc caccgatgtg cgcagtgcg ctgacagcga ctataaccgt 420
tatcacaagc gcagcgaat acagggtaac ctgacgcagc aattaggggc ctatggctct 480
gtttatttaa atttaacgca gcaggattac tggaacgacg caggtaacaa gaacacggta 540
60 tcggcggggtt acaacggacg tattggcaag gtcagttaca gtattgcata tagctggaat 600
aaaagccctg aatgggatga aagcgatcgc ttgtggtctt tcaatatctc cgttccacta 660
ggccgggcct ggagtaacta tcgcgtcacg accgacagg atggctcgta caatcaacag 720
gttgggggtca gcggaacgct gcttgaggat cgcaacctga gctacagtgt ccaggaaggc 780
tacgccagca acggtgtggg taacagcggg aacgctaacc ttggctatca ggggtgggtcc 840
65 ggtaatgtca acgtaggcta tagctacggg aaagattacc ggcagctcaa ctacagcgtt 900
cgccggcgcg tgatagttca tagcgaaggc cccaaccgct aggcgaaacc 960
atgacgctca tctccgtacc cgggtcgcgc aatgcccgcg tgggtgaataa cggcggcgctt 1020

cagggttgact ggatgggttaa cgcgcatcgtg ccttatgcc a tgccgtatcg tgaaaacgaa 1080
atctcactgc gtagcgattc gttgggtgac gatgttgacg ttgaaaatgc gttccagaaa 1140
gtgggtgccaa cgcgtggagc gattgtcaga gcgcgttttg ataccgcgt tggttacgc 1200
gtattaatga cgtctcttcg ttccgcgggc agcccggtgc cttttggagc aacggcaacg 1260
5 ctaatcaccg ataaacaaaa cgaggtgagc agtatcgttg gtgaagaagg acagctctat 1320
attagcggaa tgccagagga aggacgggta ttgattaaat ggggtaaatga cgcgtcgcag 1380
caatgcgtgg cgccttataa attatccctg gaattaaaac agggcggaat tattcctgtt 1440
tcggccaatt gccagtaa 1458
<212> Type : DNA
10 <211> Length : 1458
SequenceName : SEQ ID 420
SequenceDescription :

Sequence
15 -----
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgagtgggtt acaccgtcaa gcctcctacc ggagacagca atgagcagac acaattttatt 60
gattattttta atctgttcta cagtaagcgt gatcaggaac aaataagcat ctctcagcag 120
20 cttggaaatt acggtgcgac atttttcagt gccagtcgcc aaagtactcg gaacacgtca 180
cgcagcgacc agcaaatatc atttggatta aatgtgccgt ttggtgatat tacgacttcg 240
ctgaattaca gctattccaa taatatatgg caaaacgacg gggatcattt actcgctttt 300
acgcttaatg ttcccttcag tcattggatg cgtacagaca gtcagtcggc atttcgtaat 360
tcaaacgcca gttacagtat gtcaaacgat ttgaaaggcg gcatgaccaa tctatcgggg 420
25 gtttatggca ctctgctgcc ggataataac ctgaattata gcgttcaggt cggtaacacc 480
cacggaggtg atacatcgtc tggcaccagt ggttacagta ctcttaatta tcgtggagct 540
tacggcaata ctaatgtcgg ttacagtcgg agtggtgaca gcagccagat ttattacgga 600
atgagtgggtg ggattattgc tcattgctgat ggcacacct ttggacagcc gctgggcgac 660
acaatggttc tggttaaggc tcctggcgct gataatgtca aaatagagaa ccagaccgga 720
30 attcataccg actggcggtg ctatgccata ttaccatttg cgacagaata tagagaaaat 780
cgtgtcgttc ttaacgcgaa ttcccttgca gataatgttg aactggatga aaccgtggtc 840
actgtcatcc caactcacgg tgctattgcc agagcaacat ttaatgcaca aatcggcggg 900
aaagtattaa tgacgttgaa gtacggtaat aaaagcgttc cattcggtgc aattgtcact 960
cacggagaga ataaaaatgg cagcattgtc gcggaaaacg gtcagggtta tctgactgga 1020
35 cttccacagt cagggaaatt acaggtttca tggggcaatg ataaaaactc aaactgtatt 1080
gtcgattaca agcttcctga agtctctcct ggaaccttgc tgaaccagca gacagcaatc 1140
tgtcgctaa 1149
<212> Type : DNA
<211> Length : 1149
SequenceName : SEQ ID 421
SequenceDescription :

Sequence
45 -----
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgtctgctt tgtatgaacg ttcacagctg acgcaggtga tgatttcacg tgccccggcg 60
actgctgaaa ccatggagaa ggcggaatat ctgcgcctgg actgcacat caaggaagtc 120
cagttcacccg ccggtcagaa acaggatatt gatgtgacca cgctctgctc cacagagcag 180
50 gagaacatca acggtctggg ggcgtcgtcc gagatttcca tgtcgggtaa tttttatctg 240
aatcaggccc agaacgccct gcgtgatgcc tatgacaatg acacggtgta tgcgtttaag 300
gtgcagtttc cgtccggtaa gggctttaag ttccctggcg aagtgcgtca gcacacctgg 360
tcatccggta ccaacggcgt ggtggctgca acgttttcac ttccgctgaa gggtaaaccg 420
gtgtcctatg tggtagcgtt ggcgtttgtg aaaaatctgg ataagacact taccgtgaat 480
55 accggtgcgc tgcgtgacaat gtcagtcagt gtcaacgggg gaacgcggcc ttataaacac 540
gcctggaaga aggatggtca gccggtagag ggacagacta ctgacacttt cagtaagcca 600
ggtgcgcagt caggtgataa gggggcttat acctgcgagg taacggattc tgcagaacag 660
ccgcagagca ttacctctga tgcgtgtaca gtaacgggtta atgggtcggg cggataa 717

60 <212> Type : DNA
<211> Length : 717
SequenceName : SEQ ID 422
SequenceDescription :

65 Sequence

<213> OrganismName : Escherichia coli O157:H7

<400> PreSequenceString :
atgagaaaca aacctttttaa tcttctgtgc gcttttttgt ggctggcagt gagtcacgct 60
ttggctgcgg atagcacgat tactatccgc ggctatgtca gagataacgg ctgtagtgtg 120
gccgctgaat caaccaatatt tactgttgat ctgatggaaa acgcggcgaa gcaattttaac 180
5 aacattggcg cgacgactcc tgtcgttcca tttcgtattt tgcgtgcacc ctgtggtaac 240
gccgtttctg ccgtaaaagt tgggtttacc ggcgttgacg atagccacaa tgccaacctg 300
cttgacttg aaaatacggg gtcagcggct tcgggactgg gaatacagct tctgaatgag 360
cagcaaaatc agataccctt taatgctcca tcgtccgcga tttcgtggac gaccctgacg 420
ccgggttaaac caaatacggt gaattttttac gcccggttaa tggcgacaca ggtgcctgtc 480
10 actgcggggc atatcaatgc cacggctacc ttcactottg aatatcagta a 531

<212> Type : DNA
<211> Length : 531
SequenceName : SEQ ID 423
15 SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7

20 <400> PreSequenceString :
atgaataaat ccgttgtgtc aatttctgcg gcaatgttgg ttttactttg ccaaccggtc 60
atggggagcg aaatctcacc cgcaacaccg tcagatgaag acaactacac ctttgaccog 120
caactcttcc gcggcagcag atttagtcag tcgtcattag caaaactgac aacacgtgag 180
tctgttgac cgggcaatta taaaatggat atctacacca acaataagtt gtcaggcagt 240
25 tggaaatgtca cgtttaaaga agccgctgat ggtcgcgttc tgccctgcct gacgcctgaa 300
gtcggcgacg gtctagcgcc caaaacaggg gaagataagg gggaaaaaga tcctgtctgt 360
acgtttgcta aggaactcgc tcccggcatc accagccaga cacagttgtc acaattgcgc 420
ctggacttat cgggtccaca gactcaattg attagtcgcc ctgcgggcta tgttcccccc 480
agcgagctgg ataccggagc atcgctggcg ttcatgaatt atattgccaa ctattacaac 540
30 gttgcctatt cagggcagaa tgctcatagc cagcgttcgc tatgggcatc atttaattgt 600
ggcatcaacc ttggtgcctg gcaatatcgt cagttatcca acatgacctg ggataatgac 660
aaagggaaatc agtggacaa tattcgtagc tatttgcaac gcccgctgcc cgccataaat 720
agccagttaa tgatggggca gcttatcacc agcgggaagt ttttctctgg actcagttat 780
cacggcgcta gtctcgcgac cgatgaacgt atcgtccgg actccatgcg cggctatgcg 840
35 ccgactattc gcggcggtgg cgcaacaaac gccagagtct cggtaatgca aaacgggtcat 900
gaaatatatc agâccaccgt ggctcctggc cctttcgaga taaacgacct ataccgccac 960
agctacagcg gcgatctgga tgtcaccggt acggaagcta acggcgagc cagtcgtttc 1020
agtgctccct tttcagcggt accagaatcg accgagcag atcgtcccg gaacttcccg ttataacgtg 1080
gaagtaggta aaacgcagga tagtggtgat gactcgatgt ttggtgacct tacctggcag 1140
40 cacgggatga ctaatacgtc gacatttaac agtgggtcgc gtatcgctga tggctaccag 1200
gcgctgatgc tgggcggagc ctatggcagt tcgctggggg catttggggc aaacctcact 1260
tggtcccatg cgcgtgttcc cgaaagcgaa cgcagagtgt gttggatgtc gcaattaacc 1320
tggagtaaaa ctttccagcc tacttcaacc accgtctccc tggcagggtta tcgatactct 1380
accagcggtc atcgtgatct ggctgatgtg ctgggagagc gtcatgctgc cagcaataaa 1440
45 cagtcattggg actccagcca gtggcgctcaa cagtcgcgct tcgatcttac gtttaagtcag 1500
agccttgcta attacggcaa cctgtttgtg tcaggttcaa cacagaacta ccgtggcggc 1560
aagagccgtg atacacagct tcagttagggt tacagcaata gctttagcca tggcatcagt 1620
atgaaccttt ccgtcggacg ccaaaagaatg ggcggctata aagacaattc tgatgatatg 1680
cagacggtaa catccctttc attctcattc ccacttggcg gcaatggacc tcgtgtacca 1740
50 agtcttagca acagctggac ccattcaact gacggtagct cgcaattaca aagctcgcta 1800
accggaatgc ttgatgaagc acagaccacc aactacagcc tgaacgtcat gcgcgatcaa 1860
caatataagc agacgacgct tagcggaaac atgcaaaaaac gtttttcaca aactaccgtc 1920
ggattgaacg catcgaaggc ccaggattac tggcaggctt caggtaacgt acaaggcgcg 1980
atggctgtgc atgggtggcg cattactttc ggacccctatc tgggtgaaac gttcgcctcg 2040
55 gtcgaagcta aaggcgcgaga aggtgcaaaa gtctataact ccagtcagct ggaaattaat 2100
gacagtggct atgcgcttgt tccggcagta acgcccctatc gctacaaccg tatatctctc 2160
gatccacaag gaattggatgg cgatgccgag ttggctcgaca gtgaaagaca ggtagcaccg 2220
gttcgggtg cggcggtgaa agtaattttc cgtacccgtc ctggtaaagc gttgctgatt 2280
aaatcccgca tggcagatgg ttcggaactg ccaatgggag ccgatgtgct ggatgagaat 2340
60 aatacagtcg tcggtatagc cggtcagggg gggcaaatat acctccgcac agaacagaca 2400
aaaggccact tgtcagttcg ctggggtgaa ggtgctaacg atagctgcca attgcccttt 2460
gatatcagcg ggaaggacag caatagccct atcatccgc tgaatgaaac ctgtcagttc 2520
tga 2523

<212> Type : DNA
65 <211> Length : 2523
SequenceName : SEQ ID 424
SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
 <400> PreSequenceString :

5	atgaaactcg	cgcctgttt	tctgacactc	cttcctggct	tgcgcgttgc	cgccagctgg	60
	acttctccgg	ggttccctgc	ctttagcgaa	cagggaaacg	gaacatttgt	cagccacgcg	120
	cagttgccca	aaggtacgcg	tccactcacg	ctaaattttg	accagcagtg	ctggcagcct	180
	gcagatgcga	taaaactcaa	tcagatgctt	tccctgcaac	ctttagcaaa	cacgcccgtc	240
10	caatggcgat	tgttcaggga	cggcaaatat	acgttgcaaa	tagaoacccg	ctccggtacg	300
	ccaacattga	tgatttccat	ccagaacgcc	gccgaaccgg	tagcaaacct	ggtccgtgaa	360
	tgcccgaat	gggatggatt	accgctcacg	ctggatgtca	gcgccacttt	cccgggaagg	420
	gccgccgtac	gggattatta	cagccagcaa	attgcatag	tgaagaacgg	tcaaataacg	480
	ttacaaccgg	ctgctaccag	caacgggtta	ctcctgctgg	aacgggcaga	aactgacgcc	540
15	tctgcccctt	tgcactggca	taacgccacg	gtttactttg	tgctgacaga	tcgtttcgaa	600
	aacggcgatc	ccagtaata	ccagagttac	ggacgtcata	aagacgggat	ggcggaaatt	660
	ggcacttttc	acggcggcga	tttacgcggc	ctgaccaaca	aactggatta	cctccagcag	720
	ttgggcgtta	atgctttatg	gataagcgcc	ccatttgagc	aaattcacgg	ctgggtcggc	780
	ggcgggtaca	aaggcgattt	cccgcattat	gcctaccacg	gttattacac	acaggactgg	840
20	acgaatcttg	atgccaatat	gggcaacgaa	gccgatctac	ggacgctggg	tgatagcgca	900
	catcagcgcg	gtattcgtat	tctctttgat	gtcgtgatga	accacaccgg	ctatgccacg	960
	ctggcggtata	tgacggagta	tcagtttggc	gcgttatatc	tttctgggtg	cgaagtgaag	1020
	aaaacgctgg	gtgaacgctg	gagcgactgg	aaacctgccc	ccgggcaaac	ctggcatagc	1080
	tttaacgatt	acattaattt	cagcgacaaa	acaggctggg	ataaatgggtg	gggaaaaaac	1140
25	tggaatccgta	ccgatatcgg	cgattacgac	aatcctggat	tcgacgatct	caccatgtcg	1200
	ctagcctttt	tgccggatat	caaaaccgaa	tcaactaccg	cttctgggtc	gccgggtgtt	1260
	tataaaaaca	aaacggatac	ccacgctaaa	gccatcgacg	gctttacccc	tcgcgattac	1320
	ttaaaccact	gggttaagtc	gtgggtccgc	gactatggga	ttgatgggtt	tcgggtcgat	1380
	accgcaaac	atggtgagtt	gcccgcttgg	cagcaactga	aaaccgaagc	cagcgccgcg	1440
30	cttcgcgaat	ggaaaaaagc	taaccccgac	aaagcattag	atgacaaacc	tttctgggat	1500
	accggtgaag	cctggggcca	cggcgtgatg	caaagtgact	actatcgcca	cggcttcgat	1560
	gcgatgatca	atttcgatta	tcaggagcag	gcggcgaaag	ctgtcgattg	tattgcgcat	1620
	atggatacga	cctggcagca	aatggcggag	aaattgcagg	gtttcaacgt	gttgagctac	1680
	ctctcgtcgc	ctgatcccg	tctgttcctg	gaagggggcg	acaaagcagc	agagttatta	1740
35	ctattagcgc	caggcgcggt	acaaatcttt	tatggcgatg	aatcctcgcg	tccgttcggt	1800
	cctacaggtt	ctgatccgct	gcaaggtaca	cgttcggata	tgaactggca	ggatgttagc	1860
	ggtaaactcg	cgcgaacgt	cgcgcactgg	cagaaaatca	gccagttccg	cgcccgccat	1920
	cccgcgaatt	gcgcgggcaa	acaaacgaca	ctttcgctga	agcagggcta	cggctttggt	1980
40	cgtgagcatg	gcgacgataa	agtgcctggt	atctgggctg	ggcaacagtg	a	2031

<212> Type : DNA
 <211> Length : 2031
 SequenceName : SEQ ID 425
 SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
 <400> PreSequenceString :

50	atgccacaac	gacaccacca	gggacataaa	cgcacaccga	aacagttggc	gctcattatc	60
	aaacgctggt	tgccgatggg	gctcactggc	agcggcatgc	tttgcaactac	cgctaaccgc	120
	gaagagtatt	atttcgaccc	cattatgctg	gaaaccacaa	aaagtgggat	gcaaaacacc	180
	gatctgtcac	gtttttcaaa	aaaatacgca	caactaccag	gaacttatca	ggttgatatc	240
	tggttgaata	aaaagaaggt	ttcacagaaa	aaaattacat	ttaccgcca	tgacagcaaa	300
55	cttctgcagc	cacagtttac	ggtagaaca	ctacgtgagc	tgggtattaa	ggtggatgaa	360
	atcccggcgc	tggttgaaaa	agatgacgat	agcgtgatca	actcgttgga	acaaatcatt	420
	cccgggtacag	ctcgtgaatt	tgatttcaat	catcagcgac	ttaatttgag	cattcccaa	480
	attgcactgt	accgtgatgc	aagaggttac	gtctcccctt	ctcgttggga	cgatggtata	540
	ccaacgctgt	ttaccaacta	ctcgtttaca	ggttctgata	accgttaccg	ccagggaat	600
60	cgtagccaac	gacagtacct	aaatagcaa	aatgggtgca	atthttggcc	ctggcgatta	660
	cgtaactatt	ctacgtggac	acgcaacgat	caggcgtcaa	gctggaacac	tatcagtagt	720
	tatttacaac	gtgatatcaa	ggcgttgaa	tctcagttgc	ttctgggaga	aagcgccacc	780
	agcggcagta	ttttttccag	ctacaacttt	actggcgtgc	aactcgcttc	cgacgataat	840
	atggttgcaa	acagccagcg	cggatttgcc	ccaacggtag	gcggtatcgc	aaacagtagt	900
65	gcaatcgtga	ctatcaggca	aaatgggtat	gtgatctatc	aaagcaacgt	gccagcgggt	960
	gcctttgaaa	ttaacgatct	ctacccctct	tccaacagcg	gcgatttaga	agtcacgatt	1020
	gaagaaagtg	acggtacgca	acgtcgtctt	atccagcctt	attcttcatt	acccatgatg	1080

```

cagcgacctg ggcattctaaa atatagcgcg accgctggac gctatcgcgc tgatgcaaac 1140
agtgtatagca aggaacccga atttgctgaa gccacggcaa tatatggttt gaataatact 1200
tttacgctgt atggcgccct gctcgggtct gaagattatt atgcgctggg gatcggtatc 1260
ggcgccacac ttggcgccact gggcgcgctt tcatggctat caatggcgta cgcagtacat caaagatatt 1320
5 ttcgataacc agcactcttt tcatggctat caatggcgta cgcagtacat caaagatatt 1380
ccggaaacca acaccaatat cgctgtcagc tactatcgct ataccaacga tggctatttt 1440
agttttgatg aagccaatac ccgcaattgg gactataaca gtcgccaaaa aagtgaattt 1500
caattcaaca tcagccagac aatatttgat ggggtaagtc tgtatgcctc cggttcacag 1560
caagactatt ggggcaataa cgagaaaaac aggaatatct ctggtggggg ttccggccag 1620
10 caatggggaa ttggttacag cctgaattat caatacagcc gctacactga tcaaaataat 1680
gaccgcgcac tctctttgaa tctcagtatt ccgttagaac gctgggtacc gcgtagccgg 1740
gtttcctatc agatgaccag ccagaaagat cgcccaaccc aacatgaaat gcgtcttgat 1800
ggctcactgc tggatgatgg tcgcctgagc tatagtctgg aacaaagtct ggatgacgat 1860
aacaacataa acatgacgtt gaacggcagt taccgttcac ctatgggaa cttcagtgcc 1920
15 ggatacagtt acggtaataa cagtagccaa tacaattacg gcgttacccg cggcgtgggt 1980
atccatcctc atgggtgtgac gctctcgcaa tatctgggca acgcttttgc gcttattgat 2040
gctaaccggg catctggcgt gaggatacaa aactatccgg ggattgtac tgatcccttt 2100
ggctatgagc tgggttcctta tctcacaact taccaggaaa accgtctctc ggtagatact 2160
acgcagctgc ccgataacgt cgatcttgaa caaacaacac agttttgtgg gcccacaga 2220
20 ggtgcaatgg tagcggcgcg tttcaacgcc aatatcggtt atcgcgta ctgttacagtc 2280
agcgatcgca acggtaaaacc gttgcccttt ggcgctcttg ccagcaacga tgatacggg 2340
caacaagata tcgttgatga gggcggcata ctatatctc ctgggatac gagtaaatca 2400
caaagctgga ctgtacgctg gggaaatcag gcagatcaac aatgtcagtt tgcttttagt 2460
acaccggatt cagaaccaac aacctctgta ttacaaggca cagcgagtg ccattaa 2517
25
<212> Type : DNA
<211> Length : 2517
      SequenceName : SEQ ID 426
      SequenceDescription :
30
Sequence
-----
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
35 atgatgttca gaaatagaat attactaata tttatattgt gggctaattt tacctgggct 60
gggtgtcgta ctactgcac attaaatatt acagatggta ttaattgttg ggagatttta 120
gcgaatgaaa cttccttttag taaaagtgtc gtgtttactg ggatatcttg tgatacgagc 180
acggataaaa tagtttataa aaatatccaa agtgattggg ttgaagttgg gccttttggt 240
aatggcgaaa aattaaagg taaaatagag tctttaggta aaaccagcga cacaattggg 300
40 aaatccagca atgcgcaggg agtattacct tatgtgtgta aaatagccag aggcacacct 360
gattttactg gagaagaaaa atctacctgg tttatttcag ataccgtgat tgcaaattt 420
ggcgggtgag catcgtcac ccatcgatttt ttggtgggta ttgttaaggc attgaagttt 480
aactgggtgtg tgaattatct caccagcaaa ctggcggggg atacatttac gcttgggtta 540
aatatttctt attatcctaa aaatacgacc tgtaagcctg aaaacaccgt tataaaagta 600
45 gatgatatac ccttgttcca gtcagaaat cagggaagaa ttgcggcgaa cagtaaggaa 660
ggaacaatta cgttgaaatg tgataatctt ttccggcgaca aaaaacaagc atcgcggaat 720
atgggtgtat atctttctag cagtgtacta gttaaaggaa gtaatactat tttgcgtggg 780
aaaacagata atggtgtagg gtttgtgttg gatctaacag aaccaccaa agggactgag 840
gctgccatta aaatttcggc caacggcgat cagggcgcgg cgacatcatt atggaaaaa 900
50 gataaaccag gagtttcatt aaatagcaac attattaata taccagtcac ggccagttac 960
tatgtatatg atgaaaaaaa agttaaatct ggcgcactgg aagcaaccgc attaatcaac 1020
gtgaaatacg attaa 1035
25
<212> Type : DNA
<211> Length : 1035
55      SequenceName : SEQ ID 427
      SequenceDescription :
Sequence
-----
60 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgattaaaa aagcttcgct gctgacggcg tgttctgtca cagccttttc cgcttgggca 60
caggatacca gcccggtatc tctcgctggt actgctaacc gttttgaaca gccgcgcagc 120
actgtgcttg caccaaccac cgttgtgacg cgtcaggata tcgaccgctg gcagtcgacc 180
65 tcggttaaat atgtgctgcg ccgtcttcgg ggcgtcgata tcacccaaaa cggcggttca 240
ggtcagctct catctatttt tattcgcggt acaaatgcc gtcagtgtgt ggtgttaatt 300
gatggcgctac gcctgaatct ggcggggggg agtgggtctg ccgaccttag ccagttccct 360

```

```
attgcgcttg tccagcgtgt tgaatatatc cgtggggccac gctccgccgt ttatgggtcc 420
gatgcaatag gtgggggtggt gaatatcatc acgacgcgcg atgaaccggg aacggaaatt 480
tcagcagggt ggggaagcaa tagttatcaa aactatgatg tctctacaca gcaacaactg 540
ggggataaga cacgagtaac gttgttgggc gattatgccc atactcatgg ttatgatggt 600
5 gttgcctatg gtaataccgg aacgcaagcg cagccagata acgatgggtt tttaagtaaa 660
acgctttatg gcgcgctgga gcataacttt actgatgcct ggagcggctt tgtgcgcggc 720
tatggctatg ataaccgtac caattatgac gcgtattatt ctccgggttc accattgggtc 780
gatacccgta aactctatag tcaaagtggg gacccggggc tgcgatataa cggcgaactg 840
attaaatcac aactcattac cagctatagc catagcaaag attacaacta cgatcccat 900
10 tatggtcgct atgattcgtc ggcgacgctc gatgagatga agcaatacac cgtccagtgg 960
gcaacaaca tcatcattgg ccacggtaat gttggtgcgg gtgttgactg gcagaagcag 1020
agcacggcac cgggcacagc ttatgttaag gatggatatg atcaacgtaa taccggcatc 1080
tatctgaccg ggctgcaaca agtcggcgat tttacctttg aaggcgcagc acgcagcgac 1140
gataactcac agtttggtcg .tcatggaaac tggcaaacca gcgcgggttg ggaattcatc 1200
15 gaaggttatc gcttcattgc ttctacggg acatcttata aggcaccaa tctggggcaa 1260
ctgtatggct tctacggaaa tccgaatctg gaccgggaga aaagcaaaa gtgggaaggc 1320
gcgtttgaag gcttaaccgc tggggtgaac tggcgatttt ccggatatcg taacgatgtc 1380
agtgaactga tcgattatga tgatcacacc ctgaaatatt acaacgaagg gaaagcgcgg 1440
attaagggcg tcgaggcgac cgccaatttt gataccggac cactgacgca tactgtgagt 1500
20 tatgattatg tcatgacgcg caatgcaatt accgacacgc cgttggttacg ccgtgctaaa 1560
cagcaggtga aataccagct cgactggcag ttgtatgact tcgactgggg tattacttat 1620
cagtatctta gcaactgcta tgataaggat tactcatctt atccttatca aaccgttaaa 1680
atgggcgggtg tgagcttggt ggatcttgcg gttgcgtatc cggtcacctc tcacctgaca 1740
gttcgtggta aaatagccaa cctgttcgac aaagattatg agacagtcta tggctaccaa 1800
25 actgcaggag gggaatacac cttgtctggc agctacacct tctga 1845
<212> Type : DNA
<211> Length : 1845
SequenceName : SEQ ID 428
SequenceDescription :

30
Sequence
-----
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
35 atgaaaaaca aattgttatt tatgatgtta acaatactgg gtgcgcctgg gattgcagcc 60
gcagcagggt atgatttagc taattcagaa tataacttcg cggtaaatga attgagtaag 120
tcttcattta atcaggcagc cataattggt caagctggga ctaataatag tgctcagtt 180
ggcaggggag gctcaaaaact tttggcggtt gttgcgcaag aaggtagtag caaccgggca 240
aagattgacc agacaggaga ttataacctt gcataatatt atcaggcggg cagtgcgaat 300
40 gatgccagta tttcgcaagg tgcttatggt aatactgcga tgattatcca gaaaggttct 360
ggtaataaag caaatattac acagtatggt actcaaaaaa cggcaattgt agtgacagaga 420
cagtcgcaaa tggctattcg cgtgacacaa cgttaa 456
<212> Type : DNA
<211> Length : 456
SequenceName : SEQ ID 429
SequenceDescription :

Sequence
-----
50 <213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgaacattt ttgcatattt actgggtactt gtattttcca tgagcatgag cagcagcgcg 60
tttgccagcg tggtaatgac cggaaccgct attattttcc ctggtgacgc aaaggaaaaa 120
accatccagt tgcgaaatac cagcgatcag ccctatatca ttaatatcca tgttgaggat 180
55 gaacgtgggt ctgacaagaa tgtaccgttt atgccaaccc cgcagacatt tcgcatggaa 240
gctgccgcag gtcaggcggt acgcctgctc tacactggta ataatttacc gcaggatcgc 300
gagtcgtgtt tctgggttag tttcagtcaa ctaccttatc tgaataagaa tgataaaagt 360
cagaaccagc tcactcctggc cctgactaat cgagtcaaaa ttttctatcg tcccagctcg 420
attgtcggta aatccagtga cgcacccaaa aacctgactt accaggtaaa acagaaccgc 480
60 attgaagtga cgaatcccac gggtatttac gtcacaattc gcgcgcgtga actgcttaat 540
aatggtaaaa aagtccccct cgcgaattcg gtaattgatt ctctctaaag cacaactgaa 600
tggacactac cctctggcat cagtgtcgct cccggtgcgc agatccattt agtgaccgtc 660
aaccgactat gcgtaaatgt tacgtctgag catgccttat aa 702
<212> Type : DNA
65 <211> Length : 702
SequenceName : SEQ ID 430
SequenceDescription :
```